

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

GM protein - protein search, using sw model

Run on: February 28, 2003, 16:04:04 : Search time 50 seconds
(without alignments)
3800.308 Million cell updates/sec

Title: US-09-915-543-15

Perfect score: 7692

Sequence: 1 MHSNPKVKSFSFGNTQSSP.....ADVGQGHSGGPNQNMNF 1426

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Fried. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7457.5	97.0	1435	22	ABR11808 Human BCL9 homolog
2	2026.5	26.3	1494	23	AAU078460 Mouse beta-catenin
3	1228.5	16.0	738	23	AAU078463 Human beta-catenin
4	610.5	7.9	1685	22	ARG04839 Novel human diazo
5	610.5	7.9	1693	22	ARG15619 Novel human diazo
6	578	7.5	1196	13	AAK28916 Type III procollag
7	571	7.4	1669	23	ABR57334 Mouse ischaemic co
8	565	7.3	1064	17	AAK93254 Collagen-like poly
9	565	7.3	1064	19	AAK57652 Collagen-like poly
10	565	7.3	1065	14	AAK37741 Collagen-like poly

11	559.5	7.3	1047	22	ABR70874
12	558.5	7.3	2703	22	ABR63299
13	555.5	7.2	1838	15	ABR53257
14	553	7.2	1690	22	AAK23916
15	543.5	7.1	1466	22	ABR50291
16	543.5	7.1	1466	22	ABR90747
17	543.5	7.1	1469	22	ABG15191
18	536	7.0	1669	22	AAK40863
19	535	7.0	1669	23	ABR90760
20	533.5	6.9	1672	22	AAK39077
21	528	6.9	1078	21	AAV96125
22	528	6.9	1078	23	ABR80736
23	528	6.9	1078	23	ABR06628
24	528	6.9	1078	23	AAE16478
25	527	6.9	1078	16	AAK71704
26	524	6.8	1466	22	AAE02537
27	519.5	6.8	1963	22	ABR62819
28	519	6.7	1449	22	AAE02535
29	518.5	6.7	1735	22	AAE02535
30	515.5	6.7	837	16	AAK80320
31	515.5	6.7	837	18	AAK26353
32	515.5	6.7	837	19	AAK57659
33	515.5	6.7	837	21	AAV78290
34	513.5	6.7	837	19	AAK53530
35	512	6.7	1487	23	ABG1861
36	510	6.6	1466	22	AAE02534
37	510	6.6	1466	22	AAE02534
38	509	6.6	140	22	AAO05855
39	509	6.6	1464	22	ABR82454
40	509	6.6	1418	15	ABR90764
41	508.5	6.6	1418	15	AAK57651
42	508.5	6.6	1418	22	AAK57652
43	507	6.6	2663	22	ABG17147
44	506.5	6.6	1463	22	AAE02532
45	506.5	6.6	1745	22	AAK77793

ALIGNMENTS

RESULT 1

ID ABR11808 standard: peptide: 1435 AA.

XX ABR11808:

DT 11-JAN-2002 (first entry)

XX Human BCL9 homologue, SFU ID NO:2178.

XX

XX Human: cytokine; cell proliferation; cell differentiation; growth factor;

KW haematopoiesis regulation; tissue growth; immunomodulatory; activin;

KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;

KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;

KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;

KW chronic inflammatory condition; proliferative retinopathy;

KW atherosclerosis; coronary heart disease; arterial ischaemia;

KW bone disorder; osteoporosis; vascular growth disorder;

KW tissue regeneration; wound healing; infection; immune disorder;

KW cell culture; drug screening; gene therapy; antiinflammatory;

KW antisthmatic; antiarthritis; haemostatic; antiarteriosclerotic;

KW cytoskeletal; osteopathic; vasotropic; cardiact; vitreous; antibacterial;

KW antifungal; vulvectomy; antitumor.

XX Homo sapiens.

OS W0200157188-AA.

PN 09-AUG-2001.

PD 05-FEB-2001; 2001WO-0803800.

PF 03-FEB-2000; 2000US-0496914.

Protophila melanog
Drosophila melanog
Human collagen (TY
Human EST encoded
Collagen type III
Human Tumour Endot
Novel human diazo
Human polypeptide
Human Tumour Endot
Human polypeptide
Collagen type III
Collagen type III
Amino acid sequen
Human collagen dif
Collagen alpha 1 (I)
Porcine alpha 1(I)
Drosophila melanog
Collagen alpha 1(I)
Murine pro-alpha 1
CIP 3.7 protein mu
Collagen-like from
Collagen-like poly
CIP 3.7 amino acid
Amino acid sequen
Prostate cancer-as
Bovine alpha 1(III)
Bovine alpha 1(III)
Human polypeptide
Human pro-alpha 1
Human Tumour Endot
Type II collagen
Human type II coll
Novel human diazo
Bovine alpha 1(I)
Human pro-alpha 1

27-APR-2000; 2000US-0560875.
 (HSEB-) HSEB INT.
 Tang YT, Liu C, Dimaide RT;
 MPL: 2001-457740/45.
 N-PSDB: ABA09052.
 Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer -

Claim 20: Page 256-257; 196 tpp: English.

Sequences ABA10941-ABA12430 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of identifying compounds which or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoietic regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities, chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g. asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g. osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g. of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 1435 AA:

Query March 97.08; Score 7457.5; DB 42; Length 1435;
 Best local similarity 99.48; Pired. No. 0;
 Matches 1365; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

1 MHSNPFVRSPPSNTGSSPKSKGEVAVRPPTVMSPSGNTVLDSKSNCKGQSSASQSV 60
 41 MHSNPFVRSPPSNTGSSPKSKGEVAVRPPTVMSPSGNTVLDSKSNCKGQSSASQSV 100
 61 PPSGDSKSGHTTKALPQSSKGLKNCACGAKGKREKSIASDSFVLRDUPGTHMS 120
 101 PPSGDSKSGHTTKALPQSSKGLKNCACGAKGKREKSIASDSFVLRDUPGTHMS 160
 121 DIKENSADHKKSDSOHTPHSMTPSNATAPRSSTPSHGOTATPEPTPAKTKVYVF 180
 161 DIKENSADHKKSDSOHTPHSMTPSNATAPRSSTPSHGOTATPEPTPAKTKVYVF 240
 STEMANAAEAVLKGQVETIVSFHTQNSNNKTERSTAPLNTQISALRNPCKPLPQCPFA 240

DB 221 STEMANAAEAVLKGQVETIVSFHTQNSNNKTERSTAPLNTQISALRNPCKPLPQCPFA 280
 QY 241 PANQDQSSQNTLQPTPTIPAPAKPAAPRPLDRSPGVENKLHSVGSFASSTPLP 300
 DB 281 PANQDQSSQNTLQPTPTIPAPAKPAAPRPLDRSPGVENKLHSVGSFASSTPLP 340
 QY 301 DGTGNSPPNNRAVTPVSGSSSSADPKAPPVPSGSEPTLGENPDGJOLEHRE 360
 DB 341 DGTGNSPPNNRAVTPVSGSSSSADPKAPPVPSGSEPTLGENPDGJOLEHRE 400
 QY 361 RSLQTLRDIQMLPDEKEFTGAUSGGQNPGLVDGPKKPEPQLAMMAQSSLSGK 420
 DB 401 RSLQTLRDIQMLPDEKEFTGAUSGGQNPGLVDGPKKPEPQLAMMAQSSLSGK 460
 QY 421 GPKTDVGAPFGQHRVVPSPDENVPSSMSQSTIGPDHLMKTPQLAMLLQDEFT 480
 DB 461 GPKTDVGAPFGQHRVVPSPDENVPSSMSQSTIGPDHLMKTPQLAMLLQDEFT 520
 QY 481 EKKRKQGVVQVQSSLDQMMVHGGPRGVYKGPPEYQMTPELMAHGSTEPSSGIM 540
 DB 521 EKKRKQGVVQVQSSLDQMMVHGGPRGVYKGPPEYQMTPELMAHGSTEPSSGIM 580
 QY 541 PHSLEPRGAPRPNMPSQMLPFGACMINSEGEVNVNLSRQDSGVSMPLVLRK 600
 DB 581 PHSLEPRGAPRPNMPSQMLPFGACMINSEGEVNVNLSRQDSGVSMPLVLRK 640
 QY 601 DGRNPPGQGTSGRGGRFPNPGQSLSEEMVQDLAKKGLGLPRMAMEBIRSMENK 660
 DB 641 DGRNPPGQGTSGRGGRFPNPGQSLSEEMVQDLAKKGLGLPRMAMEBIRSMENK 700
 QY 661 MIPGSRHMEPPNNPFRIPREPIPLSPSGIPFGKIPQMGKRELEFEEVWPSMKKIV 720
 DB 701 MIPGSRHMEPPNNPFRIPREPIPLSPSGIPFGKIPQMGKRELEFEEVWPSMKKIV 760
 QY 721 NUNVNGSSQMLPKMKPAGAPEDMKLRGSGDMLPAQKMYPLPGEHPOVETMG 780
 DB 761 NUNVNGSSQMLPKMKPAGAPEDMKLRGSGDMLPAQKMYPLPGEHPOVETMG 820
 QY 781 PPEPLPMGSGPSSNSGLRMLREPIGPDQRTNSRLSHMPLPLNPSNITSLNTAPVGR 840
 DB 821 PPEPLPMGSGPSSNSGLRMLREPIGPDQRTNSRLSHMPLPLNPSNITSLNTAPVGR 880
 QY 841 LGRKPLDTSVAGSVHSPINFLKSPTHOVQSPMLGSPGSLKSPVPSLALAGA 900
 DB 881 LGRKPLDTSVAGSVHSPINFLKSPTHOVQSPMLGSPGSLKSPVPSLALAGA 940
 QY 901 AASTKSPVYLGSAAAPVHLKSPSLAPSPGWTSSPKPLPQSPQIPNKAALIMASHA 960
 DB 941 AASTKSPVYLGSAAAPVHLKSPSLAPSPGWTSSPKPLPQSPQIPNKAALIMASHA 1000
 QY 961 MGVNVSQGPPTPTASQASVNIHGLSPSTPYTPMPPLTSLQNTSLIMSKSKSFAMS 1020
 DB 1001 MGVNVSQGPPTPTASQASVNIHGLSPSTPYTPMPPLTSLQNTSLIMSKSKSFAMS 1060
 QY 1021 SFPLY-HDAIKIVASSDQSPFASRSLNPLPSNNKMGINTNBRISLNPVPMRLTSP 1079
 DB 1061 SFPLY-HDAIKIVASSDQSPFASRSLNPLPSNNKMGINTNBRISLNPVPMRLTSP 1120
 QY 1140 GPPVQSPQVQVFPNPGSGGSGSPGQMGPGGPGFGLRPNLQSSAAALDFKQCG 1240
 DB 1181 GPPVQSPQVQVFPNPGSGGSGSPGQMGPGGPGFGLRPNLQSSAAALDFKQCG 1280
 QY 1200 GPDSTVLNCSMPVFTDPLQVIRKAGTIPPEFLSKLISEKRSQTLQVFGKEVIG 1279
 DB 1241 GPDSTVLNCSMPVFTDPLQVIRKAGTIPPEFLSKLISEKRSQTLQVFGKEVIG 1300
 QY 1260 KQPGQGPQGFSHMGQMGQAFRMGLALPQMGQGPVSTPPLQSLAPSMQGHNMKPP 1319
 DB 1301 KQPGQGPQGFSHMGQMGQAFRMGLALPQMGQGPVSTPPLQSLAPSMQGHNMKPP 1360

QY 1320 AFIQQGMMGPHHMMSPAGSTMGPTLMSNPAAVGIMCKRQPAQIYTHPGVSPG 1379
 DB 1361 AFIQQGMMGPHHMMSPAGSTMGPTLMSNPAAVGIMCKRQPAQIYTHPGVSPG 1420
 QY 1380 MMSMGMGMGPGPO 1392
 DB 1421 MMSMGMGMGPGPN 1433

RESULT 2

AA078460
 ID AA078460 standard: Protein: 1494 AA.

AC AA078460;

DT 02-JUL-2002 (first entry)

DE Mouse beta-catenin nuclear localised protein.

XX Mouse; beta-catenin nuclear localised protein; cancer;

XX gene therapy; EST; expressed sequence tag.

XX Mus musculus.

XX W0200224738-A1.

XX 28-MAR-2002.

XX 19-SEP-2001: 2001WO-JP08140.

XX 22-SEP-2000: 2000JP-0287876.

XX (KYO) KYOKA HAKKO KOYO KK.

XX Akiyama T, Adachi S;

XX WPI: 2002-330014/36.

XX N-PSDB: ABK47631.

XX New beta-catenin nuclear localised protein for diagnosis and treatment

XX of diseases associated with nuclear localisation of beta-catenin e.g.

XX cancer.

XX Claim 1: Page 81-88; 113pp; Japanese.

XX Sequence 1494 AA:

Query Match 26.3%; Score 2026.5; DB 23; Length 1494;

Best local similarity 36.8%; Pred. No. 2.8e-110;

Matches 602; Conservative 178; Mismatches 491; Indels 463; Gaps 94;

QY 5 NIKVRSSPGNTQSSPKSKOEVNVRPPTVMSFGNPDLSKFSNKGKQASQSPS-- 62
 DB 12 HPRRREAP-GSPPLSPRGH----CPAPAPAKPM-HP--ENKLTNNGKTNCGAQSHQNV 62
 QY 63 ----PCL--SK---SGGTRKALPGRGSGMGLN-GAG----NGAKKGKREKSLATSE 108
 DB 63 NCGPTNLGSKGVAGSGHAKANQISPSNSSLKNQAGVSPSSLSKQVREKRSVDSG 122
 QY 109 DGRDPTFMDSDSIKE-----CNSA 128
 DB 123 EQREATGPELDSAPKAVARSKRRCVLEKQPSYSGFEMSGTGFSEDEKPTAAAHNCVA 182
 QY 129 DHKSDQSHTHSKTP-----SNATAPRSSTPSHGOTTATP-----T 167

DB 183 D-----PAMVTPQIGQTAQLPLSESSAV---GVH-RVQNLRIAYGSSGIV 229
 QY 168 PAKGTAKVYVEETEMNKAEEVAKGVETIVSFHUNISNNKTEPSTANILSLA 227
 DB 230 PG-KPPSGVYVFTHLANTAAEAVLQGRALSLAYHQGNPRAKLLVA----- 272
 QY 228 KNDPKPLPGQPPAPANDQSSONTRIQPTPTAPATPAAPPLPLPSPF--GVNKL 285
 DB 278 ---PKVFPPLPEPLPLN---TFSAGTQSGPPLPLPPPPAPASAPALPPLPPEL 341
 QY 286 IP-SVGSIAASTPLPDDGTGPESTNNKAVTPVSGSSSSAOLKA--PIIPVSSSEI 342
 DB 332 APTNSVG--AASGSGTGCHTPTTAATAMPPLPGQIPGASVALLGATPTGNGGN 389
 QY 343 TASENDPLSDEPLEHREKSLQTLRDIGMLP--FDEKEPT----GAQSGP-- 468
 DB 390 LVGS--EGLSKQLEHRESLQTLRDIERLLRSQETEPFKRPGGASISAPALSA 447
 QY 389 -QGNQVLDGPGKKEGPIGAMNQSGLKGPGRKTVGAPVQGNHIVYSHPEKV 447
 DB 448 ADPPSAPPGGLKTEEPLOSISQTSGL--GPLEHEVPGHVGASD-----K 495
 QY 448 PSNNSQGTIGPDHLDHMTPEQIAWLKQCFEYEEKRKRQVAVV--QSSLSGMMTHQ 496
 DB 496 QGNMNMMLGQDSL---TPGVAARKIQEYERKKEKEQIQLHGRILGLMV-- 544
 QY 507 PKGVV-KGPPRYQMTPESEMAPG-GTE--PESDGINPHSLPK-MAATPNK-NSG 544
 DB 549 MGGMGRGPPPYHKKPPDQCAPGAGALRSPMD--YDPMQLRPPPPPPRPNQMD 606
 QY 561 RLPGFAGMINSMEGPNVNPASRGLSGVSMPLDVKIPGHNPP--PQGLST 654
 DB 607 RVPFGQSGSMHEVP--MNAQGRVYRGMANNELPPIQSPFAPNAVPTG-- 659
 QY 615 PARGKRPVNGLSLSEDMFOQALAKQ-VGL--PYQAMRQILRSMENKMIINQKREH 671
 DB 660 QSEAEFETPR-VRELLRHLLEKRSKMQRPLQMGSGMSKEMKEMLV-ARHGR 712
 QY 672 GNNPFR-----RIPV-----EGPLSKRPPF-KQITVQMSGHELEPMV 714
 DB 718 A---MPQMGTMGGDLAGIPMGIEFGSGRLSTPMDQSLKREVP-KQGN----- 746
 QY 714 SKMKDVNLNVNGNSNOMIP-----QMKR-----FAGAPERMLKR-PDQSK 768
 DB 767 LNNMNVNMMNNMNLNVOMTPDQMLMSOKRQGDMMGPQGLSPEEMAVFAANSNM 824
 QY 759 PAQOKV-VPLPGEHPPQDEYGMGPPLPLPMSQGSNSGLNLEPPLGRT-- 849
 DB 827 GCPKMLMPSQPPMGQDQFSGQGPVQAMQDK-----NTWMSFQSSVPMATV 860
 QY 811 NSRLSMPTPLFPSSNPT-SLNTAPPVGRGJGRFLLASVAGVSNVSNINLKSP 869
 DB 881 TARLSM--PLPPASNPFGSYHLAS--NRGLRPPDOLITSLNQMSPNMGHKNTLS 945
 QY 870 QVQSPMLGSPSNLKSIVTBSLQSLAGLAPAA--AASIKSPVIGSA----- 941
 DB 936 QVHSPVLTSPSANLKSIPOTPOVPL--FSANPPGLKTLQVQVSSLSVQKSTG 992
 QY 922 KPSLPAFSPGWTSSPKPLQSPGIPPHKAPLTMASAMKGVNDSGPPF--IAS 977
 DB 993 KPSMAVIVSGVAVPKRAMPPGVQKKQPTLSINSSLTQNVGALPNSAINRNSVA 1052
 QY 978 PASVNTPGSLPSTYVIMPEPLISQNLSTIMSKHKKFAMPSVILYHLAKIVANS 1057
 DB 1053 PPAFSSGIMNPSLEPPTSPPTPSQNLSTLMSQMSKVPMPSTPIYNAKTIANST 1112
 QY 1038 DSEPAFSPPLPMNMPGNGINTQNPISCHNP--VIMPTLSMTHP--CUTLS 1089
 DB 1113 ELLPDK-ELLPPPPPGSG-----PGISNQDPQMKMNIANAASPPGMN-POU 1166
 QY 1090 ---NOMSPNAVPMIPPHGVPMGGLMHNPIIM---CHGQDEPVPVPGKMFPGG 1142
 DB 1167 PPTMLPFTPLGSIPIPRPNAQGTGSSQNSHMAVAGGPDLSLAPGIVPSSQMS 1226

File	1.2.3	1.2.4	1.2.5	1.2.6	1.2.7	1.2.8	1.2.9	1.2.10	1.2.11	1.2.12	1.2.13	1.2.14	1.2.15	1.2.16	1.2.17	1.2.18	1.2.19	1.2.20	1.2.21	1.2.22	1.2.23	1.2.24	1.2.25	1.2.26	1.2.27	1.2.28	1.2.29	1.2.30	1.2.31	1.2.32	1.2.33	1.2.34	1.2.35	1.2.36	1.2.37	1.2.38	1.2.39	1.2.40	1.2.41	1.2.42	1.2.43	1.2.44	1.2.45	1.2.46	1.2.47	1.2.48	1.2.49	1.2.50	1.2.51	1.2.52	1.2.53	1.2.54	1.2.55	1.2.56	1.2.57	1.2.58	1.2.59	1.2.60	1.2.61	1.2.62	1.2.63	1.2.64	1.2.65	1.2.66	1.2.67	1.2.68	1.2.69	1.2.70	1.2.71	1.2.72	1.2.73	1.2.74	1.2.75	1.2.76	1.2.77	1.2.78	1.2.79	1.2.80	1.2.81	1.2.82	1.2.83	1.2.84	1.2.85	1.2.86	1.2.87	1.2.88	1.2.89	1.2.90	1.2.91	1.2.92	1.2.93	1.2.94	1.2.95	1.2.96	1.2.97	1.2.98	1.2.99	1.3.00	1.3.01	1.3.02	1.3.03	1.3.04	1.3.05	1.3.06	1.3.07	1.3.08	1.3.09	1.3.10	1.3.11	1.3.12	1.3.13	1.3.14	1.3.15	1.3.16	1.3.17	1.3.18	1.3.19	1.3.20	1.3.21	1.3.22	1.3.23	1.3.24	1.3.25	1.3.26	1.3.27	1.3.28	1.3.29	1.3.30	1.3.31	1.3.32	1.3.33	1.3.34	1.3.35	1.3.36	1.3.37	1.3.38	1.3.39	1.3.40	1.3.41	1.3.42	1.3.43	1.3.44	1.3.45	1.3.46	1.3.47	1.3.48	1.3.49	1.3.50	1.3.51	1.3.52	1.3.53	1.3.54	1.3.55	1.3.56	1.3.57	1.3.58	1.3.59	1.3.60	1.3.61	1.3.62	1.3.63	1.3.64	1.3.65	1.3.66	1.3.67	1.3.68	1.3.69	1.3.70	1.3.71	1.3.72	1.3.73	1.3.74	1.3.75	1.3.76	1.3.77	1.3.78	1.3.79	1.3.80	1.3.81	1.3.82	1.3.83	1.3.84	1.3.85	1.3.86	1.3.87	1.3.88	1.3.89	1.3.90	1.3.91	1.3.92	1.3.93	1.3.94	1.3.95	1.3.96	1.3.97	1.3.98	1.3.99	1.4.00	1.4.01	1.4.02	1.4.03	1.4.04	1.4.05	1.4.06	1.4.07	1.4.08	1.4.09	1.4.10	1.4.11	1.4.12	1.4.13	1.4.14	1.4.15	1.4.16	1.4.17	1.4.18	1.4.19	1.4.20	1.4.21	1.4.22	1.4.23	1.4.24	1.4.25	1.4.26	1.4.27	1.4.28	1.4.29	1.4.30	1.4.31	1.4.32	1.4.33	1.4.34	1.4.35	1.4.36	1.4.37	1.4.38	1.4.39	1.4.40	1.4.41	1.4.42	1.4.43	1.4.44	1.4.45	1.4.46	1.4.47	1.4.48	1.4.49	1.4.50	1.4.51	1.4.52	1.4.53	1.4.54	1.4.55	1.4.56	1.4.57	1.4.58	1.4.59	1.4.60	1.4.61	1.4.62	1.4.63	1.4.64	1.4.65	1.4.66	1.4.67	1.4.68	1.4.69	1.4.70	1.4.71	1.4.72	1.4.73	1.4.74	1.4.75	1.4.76	1.4.77	1.4.78	1.4.79	1.4.80	1.4.81	1.4.82	1.4.83	1.4.84	1.4.85	1.4.86	1.4.87	1.4.88	1.4.89	1.4.90	1.4.91	1.4.92	1.4.93	1.4.94	1.4.95	1.4.96	1.4.97	1.4.98	1.4.99	1.5.00	1.5.01	1.5.02	1.5.03	1.5.04	1.5.05	1.5.06	1.5.07	1.5.08	1.5.09	1.5.10	1.5.11	1.5.12	1.5.13	1.5.14	1.5.15	1.5.16	1.5.17	1.5.18	1.5.19	1.5.20	1.5.21	1.5.22	1.5.23	1.5.24	1.5.25	1.5.26	1.5.27	1.5.28	1.5.29	1.5.30	1.5.31	1.5.32	1.5.33	1.5.34	1.5.35	1.5.36	1.5.37	1.5.38	1.5.39	1.5.40	1.5.41	1.5.42	1.5.43	1.5.44	1.5.45	1.5.46	1.5.47	1.5.48	1.5.49	1.5.50	1.5.51	1.5.52	1.5.53	1.5.54	1.5.55	1.5.56	1.5.57	1.5.58	1.5.59	1.5.60	1.5.61	1.5.62	1.5.63	1.5.64	1.5.65	1.5.66	1.5.67	1.5.68	1.5.69	1.5.70	1.5.71	1.5.72	1.5.73	1.
------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	----

10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
2A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
3A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
4A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
5A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80																				

[illegible]

[illegible][illegible]

[illegible][illegible]

[illegible][illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Inc.

09 protein - protein search, using sw model

Run on: February 28, 2003, 16:04:09 : Search time: 0.5 seconds

(without alignment)
4092.875 Million cell updates/sec

File: us-09-915-543-15

Perfect score: 7692

Sequence: 1 MISSNPKRVSSSTNGNTSSSP.....ANYMGPSGQGNINMM 1426

Scoring table:

BLASTM62
Gapop 10.0 , dapext 0.5

Searched: 671580 seqs, 206347115 residues

Total number of hits satisfying chosen parameters: 67158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPERMEL_21:
1: SP_archaea:
2: SP_bacteria:
3: SP_fungi:
4: SP_human:
5: SP_invertebrate:
6: SP_mammal:
7: SP_mhc:
8: SP_organelle:
9: SP_phage:
10: SP_plant:
11: SP_protist:
12: SP_virus:
13: SP_vterbrate:
14: SP_unclassified:
15: SP_virus:
16: SP_bacteriap:
17: SP_archaeap:

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Accession
1	7482	97.3	1394	4	000512
2	1588	20.6	298	11	Q9VE74
3	1584	20.6	298	11	Q90219
4	725	9.4	130	6	Q95806
5	604	7.9	1691	11	Q95802
6	559.5	7.3	1047	5	Q95813
7	558.5	7.3	2703	5	Q95827
8	557	7.2	1669	11	Q90250
9	557	7.2	1802	5	Q17163
10	554.5	7.2	1838	4	Q15094
11	552	7.2	2715	5	Q61603
12	549.5	7.1	1840	11	Q90467
13	546	7.1	1682	11	Q90289
14	540.5	7.0	1840	11	Q91103
15	535	7.0	1621	4	Q90469
16	534	6.9	1835	13	Q91A04

17	532	6.9	1447	14	Q91B91
18	532	6.9	1827	14	Q800M5
19	531	6.9	1600	4	Q90E06
20	531	6.9	1838	11	Q88207
21	530	6.9	2944	11	Q91870
22	528	6.9	1690	4	Q90M65
23	528	6.9	1691	4	Q9Y414
24	527.5	6.9	1737	11	Q90104
25	527	6.9	1691	4	Q12823
26	520.5	6.8	1485	5	Q815M4
27	519.5	6.8	1463	5	Q9VSK5
28	519.5	6.8	1466	5	Q9NMX5
29	518.5	6.7	1739	11	Q9J112
30	514	6.7	1420	13	Q90M37
31	512	6.7	1487	4	Q14047
32	511.5	6.6	1445	14	Q91251
33	509.5	6.6	1414	5	Q26334
34	507	6.6	2063	4	Q90R86
35	506.5	6.6	1160	4	Q14046
36	506.5	6.6	1745	4	Q9N206
37	506	6.6	1450	13	Q9Y104
38	506	6.6	1453	11	Q63078
39	506	6.6	1461	4	Q76045
40	506	6.6	1487	6	Q77753
41	504	6.6	2001	4	Q90B74
42	504	6.6	2062	4	Q9N179
43	503.5	6.5	1315	10	Q9SPM0
44	502.5	6.5	1442	11	Q62031
45	502.5	6.5	1442	11	Q62033

ALIGNMENTS

RESULT 1	PRELIMINARY	PR: 1394 AA.
ID 000512		
AC 000512		
DT 01-JUL-1997 (FEBRUARY, 04, (translated)		
DT 01-JUL-1997 (FEBRUARY, 04, last sequence update)		
DT 01-DEC-2001 (FEBRUARY, 19, last annotation update)		
DE 8-cell CLL/lymphoma 9,		
GN hcl9,		
US Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX NCBI-TaxId:9606;		
RN [1]		
RP SEQUENCE FROM N.A.		
RA Willis J.G., ZALTBURG L.R., Colquhoun L.J.A., Wlodarska L., STOL M.,		
RA Jaddayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,		
RA Dyer M.J.S.;		
KT "Molecular cloning of translocation t(1;14)(q21;q32) defines a novel		
KT gene (Bcl9) at chromosome 1q21.";		
RL Blood 91:1871-1873(1998).		
DR EMBL: Y13620; GAA7342.1;		
KM R-cel).		
SO SEQUENCE	1394 AA; 145970 MW; 2075012162727W; CRC64;	
Query Match	97.3% Score 7482; DB 4; Length 1394;	
Best local similarity	99.6% Prod. No. 0;	
Matches 1387; Conservative	2; Mismatches	4; Indels
	0; Gaps	0;
QY 1 MISSNPKRVSSSTNGNTSSSPKQEVNVPPTVMSTNGNOLSKPSNAGKQASASQ 60		Q91B91 xenopus lat
		Q90M65 mus sapien
DB 1 MISSNPKRVSSSTNGNTSSSPKQEVNVPPTVMSTNGNOLSKPSNAGKQASASQ 60		Q90E06 mus sapien
		Q88207 mus sapien
QY 61 PSYQSKSGSHPPKALPGVSSMGCKNAGNNAKKKKKKKSLADSPQKQKTNLNS 120		Q9Y414 mus sapien
		Q91104 rattus norv
DB 61 PSYQSKSGSHPPKALPGVSSMGCKNAGNNAKKKKKKKSLADSPQKQKTNLNS 120		Q12823 mus sapien
		Q815M4 drosophila
		Q9VSK5 drosophila
		Q9NMX5 drosophila
		Q9J112 mus sapien
		Q90M37 rattus norv
		Q14047 mus sapien
		Q91251 mus sapien
		Q26334 strimolyse
		Q90R86 mus sapien
		Q14046 mus sapien
		Q9N206 mus sapien
		Q9Y104 mus sapien
		Q63078 cytoops pyrr
		Q76045 rattus norv
		Q77753 cratis fam1
		Q90B74 mus sapien
		Q9N179 mus sapien
		Q9SPM0 mus sapien
		Q62031 mus sapien
		Q62033 mus sapien

07	108	Б	У	В	И	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	!	"	#	\$	%	&	'	()	*	+	,	-	.	:	;	<	=	>	?@	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{</
----	-----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	-----

[illegible]

RA Wang Z.-Y., Wasserman D.A., Winstock O.M., Weissbach J.S.,
 RA Williams S.M., Woodhouse T., Worley K.C., Wu D.Y., Yang S., Yao Y.-A.,
 RA Ye J., Yoon R.-E., Zaveri J.S., Zhao M., Zhang G., Zhao Y., Zhou L.,
 RA Zhou X.H., Zhou F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*."
 Science 287:2185-2195(2000).
 OR EMBL: AF004575; AAF50995.2;
 OR Flybase: FB00041617; CG15635.
 OR InterPro: IP0002052; N6_Mtase.
 OR InterPro: IP0002965; P-rich_extensn.
 OR PRINTS: PR01217; PRICHX1ENSN.
 OR PROSITE: PS00092; N6_Mtase; GNKNWN.1.
 OR SOURCE: 1047 AA; 10615 MW; 76370:AAALD736A2 (Pro-4).
 Query Match 7 18; score 559.5; DB 5; length 147;
 Post Local Similarity 24.78; Prot. No. 410-24;
 Matches 328; Conservative 116; Mismatches 448; Indels 455; Gaps 84;
 UY 198 ETVSHIHLNLSN-KTERSLAP;NTQISALRNDKRIPLQGPAP---AMQANSSNLT 254
 DB 12 ESYLPFHFVYVPPNPPLPSAPAP-----PPKRPPEPPPPPPPTLPPPIITPT 61
 UY 254 LGPFPPLPAAPKPAAPRPRLDRESVENKL-----LPSVGSIA-----SSTPPTD 601
 DB 62 TTPPTPTTTPPPPSAPPP-IGATPGVWYPLPYANAPQDPPOSHLSNSKRPDP 119
 UY 402 GTGHNSTNNKAVPPAS-----GQSSSSAAPPKAPPPPSSSGPPPTGPNKLSQQL 356
 DB 120 -GQMPFNSTVATPAPAGTPYVPPVPPQGNPPSQPTPNSLTPVAVAMPVPPVPI 177
 UY 357 EHKRSTLTLDLQRLPLPQKKEFTGAQSGDQGNFVLDPPV---KKRPPLVAMMAVSQ 414
 DB 178 GH-----PPESQ-----APTNSGVPPPALMNPDPAP-----IN 206
 UY 415 SLKGPGRITDVAPFGPQGHVPPSPDEKVPSSMSQSTTGRILLMTPLQIMLK 474
 DB 207 PDSNPPPSQTPPNSGPPV-----PPALMSPS---VPTSHDPPPSGA- 247
 UY 475 LQGFVEERKRQVNVVQCSLDIMMVHOGHPGVKCP-----PPVMTSHKMAP 529
 DB 248 -----PPNSGVPLATCTPFLQGNLTQILN 276
 UY 530 GTEFPNGLNPPHS--LPPQ--MAFHPPNSQMTPEFAGINSEMGIVNPPVASK 584
 DB 276 SHKPPNSQTPPSQVPPPSGLSQGHP-PPSQ--LPPNST-----FQSHPPTR 324
 UY 585 PGLSTVWPIPIVPRPLQKRPFGGGTFSGPQKRPFRPPVQGLSEMDQGLAKQLAP 644
 DB 425 P-----PPGLQD-PPGSMPPSLQIYP-PSMG-----WIP 352
 UY 645 P-GMAEGEIKSMEMNNMIPGSGHMFENNNPIEP-KHPEVPLSOS-KGLPKT--LPP 639
 DB 453 PDSGLLQGHPP-----PQSS--LTPPSVPLAPGSP--PLKPSNCGMHPPSILPP 398
 UY 700 QMGGRPLP-----GWPSSMKSGDVNINVMNSNSMLPP 735
 DB 399 NPAFLRLDQILKLDHGDQCLLMWYTHPPVAGFPDS---PLLPDHPVQSLPLP 433
 UY 736 KMEFAVAGPPLMKAT---GSMMLPAQKRYPLPFGHDDQVYMGKRPVPSQA- 790
 DB 454 ---NMLPPSSTPLPKNQGHPP--PQSK--PPVGLYPPST-PMIP-PLSLQGGH 504
 UY 791 ---NSGLNLEPITG--PQKRNLSHMP-PIPLNPPSNLSINAI PVG 838
 DB 505 PPSGLPLPNTGL-----PPSSTPLPNSDPLGHPNSQKPPNGLYVSTQVPPSPPT 559
 UY 839 RLQAKKPLDLSVAVSVHNSINMLKSPIMHGVSPMLASPSNLKSPQTESLAMEAD 898
 DB 560 QG-CHFGSLPNTGLPPSGSLPPSPN-----QIGHPPPSQGPPN--VAVY--604
 UY 899 PAAASAKSPVAVASAVVHVKSLIADPSQGLNSKPPQSTPPTVPPNKAT-----964

DB 605 PPSTQWIPPSGLTQSGPPSSTLPPNTGLD-PPSTLPPKPPNQSILPNTGLPSSSTP 664
 UY 955 IMASPMILGVNSGPPPLPAPASVANIQSLPSTPYIMPEPPLSQDPLSTMSKMS 1014
 DB 664 PLSGPNLIGHPQSGKPPNSITYPST--GSITPSTLQGAAP--KSL-----709
 UY 1015 KPMASSTPLVHDAIKTVASSIDSPSPKSPNLPSSNMNMLNINQNPISAPNVIM 1074
 DB 710 ---LPSNGLPSSSTL-----PPASGN--QIGHPPG--SQKPPNSG---LYP 747
 UY 1075 PT--LSIMMIGVTHSHNOMPSTNAVGN--TPHPPV-PMG-PKLSMNI-MHNSQDP 1128
 DB 748 STWMLPPSG--PLTQGHPPG-SLPPNTGLPSSSPASSTNQLGHP--GSKKP 709
 UY 1129 MPPGRRGFTQGGPPVPS--PPGVPPRPNQPSGQ-----NSPQPMPE-PGPHLQAP 1180
 DB 800 NV-----GI--YPTSGTLPKPPNQGHPQPSQKPPVNVGIVYVSTWIPSSGLQ 851
 UY 1181 SNLQSSAQAALQKPPQGGPDSFTVLAGNSPSEFTIDQLQEVINAGATLPPEDSK 1240
 DB 852 GHP-----GSGVPS-----NSVLTNS-----870
 UY 1241 PSKPSQTLQYERKPY-----PKKQPGGHPGSHQGMQGLQALQMGALQMGAL 1294
 DB 871 PPLPKNQGHPPHVSQPPNTGLPQSLPPLGSPNQLGHP--SQKPPNSITYPSTGS 928
 UY 1294 -----PGVGTID--PLGLAISMGHNMIPRALQGGKMTTHRRMSIA 1308
 DB 929 LPSGLPLQGHPPGSLDPPNTGLPSSIDP-----PLKPP--NQRNPPSQV--149 977
 UY 1339 SIMP-----GQPLMSNVAAGMIPKQKQPACTHGGVQSHQMSMGQMGQ 1391
 DB 978 SVLPDPSHPPLGSPPLQGVAS-----QKPASSKIPPTTSPQNTVW-----1K 1024
 UY 1392 QNIMPP 1398
 DB 1025 QPVKNPP 1031
 RESULT 7
 QVVEG7 PRELIMINARY: PRT: 2703 AA.
 AC QVVEG7:
 DT 01-MAY-2000 (TEMPREL: 13, Created)
 DT 03-MAY-2000 (TEMPREL: 13, Last sequence update)
 DT 01-JUN-2002 (TEMPREL: 21, Last annotation update)
 DE USA protein.
 GN OSA OR CG7467.
 OS *Drosophila melanogaster* (Fruit Fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Ephydroidea; *Drosophila*; *Drosophila*.
 ON NCBI_TaxID=7227;
 RX MEDLINE: 20196006; PubMed: 10741132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Goryunov J.D.,
 RA Amanatides P.G., Schorer S.E., Li F.W., Hoskins R.A., Gail R.F.,
 RA Goyette R.A., Lewis S.E., Richards S., Ashburner M., Johnston S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Bratton R.C., Rogers Y.H.C., Alzari R.G., Chapple M., Trotter R.D.,
 RA Man K.L., Doyle C., Baker E.G., Holt G., Nelson C.R., McKus G.L.,
 RA Abrait J.F., Atabayani A., An H.-J., Andrews-Pfankuch C., Balkwin D.,
 RA Haller R.M., Hsu A., Bakendale J., Bayraktaroglu I., Beasley E.M.,
 RA Hoesen K.Y., Hous P.V., Beyman B.P., Blandini J., Brashkov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam J.A., Butler H., Cadieu E., Cento A., Chandra L.,
 RA Cherry J.M., Cowley S., Dahlke C., Davernport L.H., Davies L.,
 RA de Fabros R., Dehner A., Deng Z., Mays A.D., Iwata D., Dietz S.M.,
 RA Dedson K., Domp L.F., Downes M., Dugan-Rocha S., Jambry H.C., Jann P.,
 RA Durbin K.D., Evangelista C.G., Foray C., Fortera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gang N.S., Gallant W.M., Glasser K.,

[illegible][illegible]

[illegible][illegible]

[illegible][illegible]

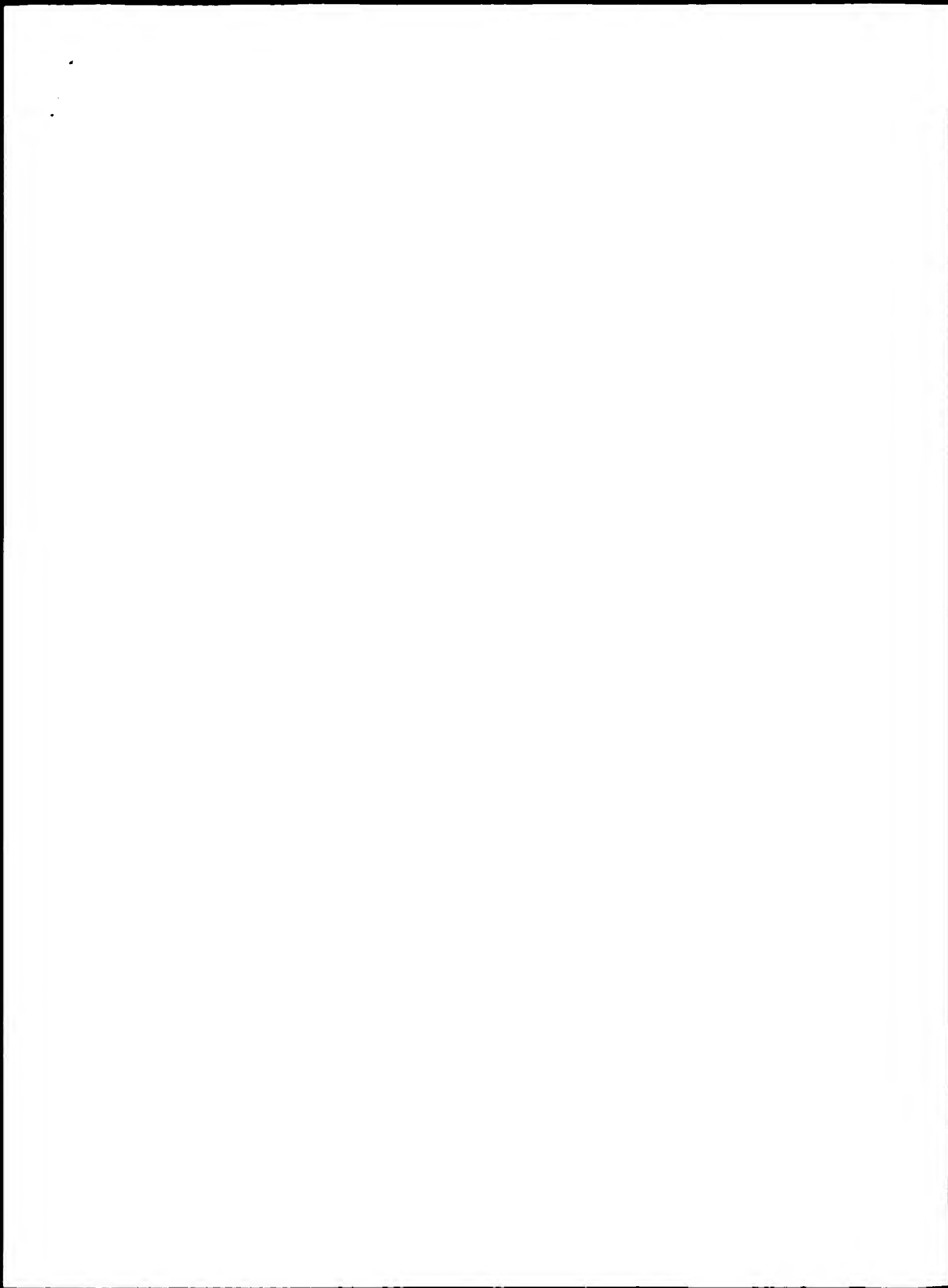
[illegible][illegible]

```

Db 1085 GEAGLEGTGCL-----EGPAQOKGEPGSLNLIWASAEKVEPPIPKKPPPI 1180
QY 1111 PLTMSNPIMH--GSGEPMPVP--GGHMF---EGGPPVQSPVQVFTTHNQS 1159
Db 1141 PGAKDURKSKGVNPPGLASPGITGSKGEGGPMPPGPDQPGIAGSPHAI EATK 1167
QY 1160 GAGDSFPGMGHPG-EGTUGSPSNIPGSSADAALPKGGRGGRPPSPVLAQNSMSVPTOP 1216
Db 1188 GDRGP-OGDPGIPGIPGIMG PGLP--GIDGVKGRKGNIGMPAN--GVVGRKTRP 1238
QY 1219 DLGVTIKWATGIEPPDLKTIIPSEKPSQTLQYTRKGVVTRKOPVGRGSPSHMGMG 1278
Db 1249 GFGGM--PGTGSPPG----ITGSKGDMG---PPG-VNGPDGPKGL-EGLSAIKRIQD 1284
QY 1279 EQAPRMGLALPGMGPPGVPVTPDIPLGTAPEMPPHNPVPPVPLQDMMTPHHGMSNAG 1338
Db 1285 DQGVPGAKGLP--GVPGPPGYDIIKG-EGGLPG--EGGPGI--KRIQD----- 1327
QY 1339 STMCGQPLTMSNPAAAVGMIPKRGHAGLYTHPPVWSPGM--MMSMGGMGPGQNMIM 1396
Db 1328 --LPG-----PKGGGVGLGVGTGTPGPTPGFDGAPGKGRMIP --- 1364
QY 1397 PPGMRKRMADVGMGSPSGGPNPG 1422
Db 1365 AGPIGPKRPPGPPGPGDIPMSMPPG 1390

```

Search completed: February 28, 2003, 16:08:41
 Job time : 112 secs



GenInfo version 5.1.4
Copyright (c) 1993 - 2003 Compugen Ltd.

00M protein - protein search, using sw model

Run on: February 28, 2003, 16:04:04 : Search time: 17 seconds

(with boot of 1000000s)
3479,342 Million (x11) updates/sec

File: US-09-915-543-15

Perfect score: 7692

Sequence: 1 MISSMPPKXKSSSSSSNCTGSSSP.....ADVMMKSSGMMNINMMI 1426

Scoring table: BLOSUM62

Gapop: 10.0 : Gapext: 0.5

Search: 112892 seps, 41476328 residues

Total number of hits satisfying chosen parameters: 112932

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match: 08

Maximum Match: 1108

Listing first 47 summaries

Database: SwissProt_40*

Prod. No. is the number of residues predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	POSTED JOB
1	605.5	7.9	1685	1 CA54_HUMAN	P27400 homo sapien
2	600	7.8	1763	1 CA24_ASCTO	P27400 ascaris son
3	571	7.4	1669	1 CA14_MOUSE	P02496 mus musculu
4	565.5	7.4	1670	1 CA54_HUMAN	P01955 homo sapien
5	559.5	7.2	1838	1 CA15_HUMAN	P20008 homo sapien
6	552	7.2	1690	1 CA44_HUMAN	P54427 homo sapien
7	543.5	7.1	1466	1 CA13_HUMAN	P02461 homo sapien
8	545	7.0	1669	1 CA14_HUMAN	P02462 homo sapien
9	530	6.9	2944	1 CA17_HUMAN	P02468 homo sapien
10	528	6.9	1758	1 CA24_CABEL	P17141 consoliated
11	523	6.8	1678	1 CA64_HUMAN	P17141 homo sapien
12	517.5	6.7	1464	1 CA13_MOUSE	P01821 mus musculu
13	516	6.7	1453	1 CA11_CHICK	P04257 gallus gall
14	506.5	6.6	1650	1 CA28_MOUSE	P04257 mus musculu
15	505	6.6	1464	1 CA11_HUMAN	P02462 homo sapien
16	504.5	6.6	1418	1 CA12_HUMAN	P02468 homo sapien
17	503	6.5	1459	1 CA18_MOUSE	P04215 mus musculu
18	502.5	6.5	1459	1 CA12_MOUSE	P04215 mus musculu
19	501.5	6.5	1453	1 CA11_MOUSE	P04215 mus musculu
20	501	6.5	1265	1 CA28_HUMAN	P13387 mus musculu
21	500.5	6.5	1603	1 CA28_HUMAN	P13387 mus musculu
22	498	6.5	1460	1 CA11_CANFA	P02492 homo sapien
23	496	6.4	1496	1 CA25_HUMAN	P08817 canis famil
24	490	6.4	1439	1 CA13_BOVIN	P05097 homo sapien
25	488	6.3	1262	1 CA13_CHICK	P12105 gallus gall
26	488	6.3	1806	1 CA18_HUMAN	P12109 homo sapien
27	484.5	6.3	1712	1 CA24_HUMAN	P08572 homo sapien
28	471.5	6.1	1707	1 CA24_MOUSE	P08572 mus musculu
29	463	6.0	1516	1 CA11_HUMAN	P02461 homo sapien
30	460.5	6.0	1464	1 CA21_BOVIN	P02465 homo sapien
31	455	5.9	1462	1 CA21_CHICK	P02467 gallus gall
32	449	5.8	1798	1 CA14_CABEL	P17141 consoliated
33	449	5.8	1775	1 CA14_MOUSE	P04257 mus musculu

44	448	5.8	1772	1 CA21_RAI	P02466 ratius norv
45	444	5.8	2142	1 BA12_HUMAN	P08664 homo sapien
46	442.5	5.8	1466	1 CA21_CANFA	P01692 canis famil
47	442	5.6	1466	1 CA21_HUMAN	P08124 homo sapien
48	428	5.5	2167	1 SER1_RAT	P08648 ratius norv
49	425.5	5.5	1455	1 CA21_RANFA	P02460 homo sapien
40	425.5	5.5	1902	1 SHP1_HUMAN	P04497 homo sapien
41	424	5.5	1472	1 CA21_MOUSE	P01149 mus musculu
42	423.5	5.5	1527	1 CA18_MOUSE	P04063 mus musculu
43	412	5.4	1779	1 CA11_BOVIN	P02464 mus musculu
44	410	5.3	1456	1 CA21_ORNOR	P02464 mus musculu
45	410	5.3	2161	1 SHR1_HUMAN	P05666 homo sapien

ALIGNMENTS

RESULT 1

ID	CA54_HUMAN	STANDARD	PRO: 1685 AA.
AC	P29400: Q16126: Q16006:		
DT	01-08-1992 (pol: 24: created)		
DT	01-08-1994 (pol: 28: last sequence update)		
DT	15-08-2002 (pol: 41: last annotation update)		
DE	Collagen alpha 5(IV) chain precursor.		
GN	COL4A5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eultheria; Primates; Catarrhini; Hominoidea; Homin;		
OX	NCBI-TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 94165049; PubMed 4120014;		
RA	Zhou J., Leinonen A., Trygvaeson K.;		
RT	"Structure of the human type IV collagen COL4A5 gene".		
RI	J. Biol. Chem. 269:6608-6614(1994).		
RN	[2]		
RP	SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS "VS (2).		
RG	TISSUE: Kidney.		
RX	MEDLINE: 92316923; PubMed 1452287;		
RA	Zhou J., Herriz J.M., Leinonen A., Trygvaeson K.;		
RT	"Complete amino acid sequence of the human alpha 5 (IV) collagen		
RI	chain and identification of a single base mutation in exon 28		
RT	converting glycine 521 in the collagenous domain to cysteine in an		
RI	Alport syndrome patient".		
RN	J. Biol. Chem. 267:12475-12481(1992).		
RN	[3]		
RP	SEQUENCE OF 85-1685 FROM N.A.		
RG	TISSUE: Placenta;		
RX	MEDLINE: 90337990; PubMed 2380186;		
RA	Pihlajaniemi T., Pohlman E.R., Myers J.C.;		
RT	"Complete primary structure of the triple-helical region and the		
RI	carboxyl-terminal domain of a new type IV collagen chain, alpha		
RT	5(IV)".		
RI	J. Biol. Chem. 265:13758-13766(1990).		
RN	[4]		
RP	SEQUENCE OF 924-1685 FROM N.A.		
RX	MEDLINE: 91169491; PubMed 2004755;		
RA	Zhou J., Hostikka S.L., Chow E.T., Trygvaeson K.;		
RT	"Characterization of the 3' half of the human type IV collagen alpha		
RI	5 gene that is affected in the Alport syndrome".		
RI	Genomics 9:1-9(1991).		
RN	[5]		
RP	SEQUENCE OF 914-1685 FROM N.A.		
RX	MEDLINE: 90160375; PubMed 1689491;		
RA	Hostikka S.L., Eddy R.L., Byers M.G., Hoeyllyar M., Shors J.B.;		
RT	Trygvaeson K.;		
RT	"Identification of a distinct type IV collagen alpha chain with		
RI	restricted kidney distribution and assignment of its gene to the		
RI	locus of X chromosome linked Alport syndrome".		
RI	Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).		
RN	[6]		
RP	SEQUENCE OF 1442-1471 FROM N.A.		

DB 1415 GIPGIPKAKGEGISLCK RGNIGIPKIDPAIPKPKGSGHPIGIPKADANL 1374
 UY 1398 PQMRPR---GMAAVGMSGSGPANG 1422
 DB 1374 POLKGEPIPIPEPKGKGFETGPGGPIG 1402

RESULT 3
 CA14_MOUSE STANDARD: PRT: 1669 AA.
 ID CA14_MOUSE
 AC 02463;
 DT 21-JUL-1986 (Rel. 01, Created)
 DI 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COLA1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
 NX MGI:10090;
 RN 11;
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89197932; PubMed=2703490;
 RA Muthukumar G., Blumberg B., Kirkien M.;
 RT "The complete primary structure for the alpha 1-chain of mouse
 collagen IV. Differential evolution of collagen IV domains.";
 RL J. Biol. Chem. 264:6310-6317(1989).
 RN 12;
 RP SEQUENCE OF 1-1154 FROM N.A.
 RA MEDLINE=88112221; PubMed=338568;
 RA Wood E., Thieriault N., Vogel J.G.;
 RT "cDNA clones completing the nucleotide and derived amino acid
 sequence of the alpha 1 chain of basement membrane (type IV) collagen
 from mouse.";
 RL FEBS Lett. 227:5-8(1988).
 RN 13;
 RP SEQUENCE OF 1149-1424 FROM N.A.
 RA MEDLINE=86301886; PubMed=375692;
 RA Nath P., Laurent M., Horn E., Sobol M.E., Zon G., Vogel J.G.;
 RT "Isolation of an alpha 1 type IV collagen cDNA clone using a
 synthetic oligonucleotide.";
 RL Gene 43:401-404(1986).
 RN 14;
 RP SEQUENCE OF 1276-1669 FROM N.A.
 RA MEDLINE=85127033; PubMed=2578561;
 RA Oberbauer L., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
 RA Vogel J.G., Voss T., Stehbold B., Glauville R.W., Kuhn K.;
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
 the alpha 1(IV) chain of basement membrane collagen as derived from
 complementary DNA.";
 RL Eur. J. Biochem. 147:217-224(1985).
 RN 15;
 RP SEQUENCE OF 1441-1669 FROM N.A.
 RA MEDLINE=87250440; PubMed=3597383;
 RA Kirkien M., Condon M.R., Blumberg B., Barlow D., Gutierrez S.,
 RA Saus J., Pithalantani F.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse
 alpha 1(IV) and alpha 2(IV) collagen.";
 RL J. Biol. Chem. 262:8436-8439(1987).
 RN 16;
 RP PARTIAL SEQUENCE FROM N.A.
 RA MEDLINE=86196099; PubMed=3009486;
 RA Sakurai Y., Sullivan M., Yamada Y.;
 RT "Alpha 1 type IV collagen gene evolved differently from the alpha
 2 collagen genes.";
 RL J. Biol. Chem. 261:6654-6657(1986).
 RN 17;
 RP SEQUENCE OF 1-28 FROM N.A.
 RA MEDLINE=89066738; PubMed=3198626;
 RA Kayes P., Wood E., Thieriault N., Kirkien M., Vogel J.G.;
 RT "Head-to-head arrangement of multiple type IV collagen genes.";
 RL J. Biol. Chem. 263:19274-19277(1988).
 RN 18;

RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89071759; PubMed=3200851;
 RA Burholo P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
 RT bidirectional promoter and a shared enhancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
 RN 19;
 RP SEQUENCE OF 1-129 FROM N.A.
 RX MEDLINE=88243724; PubMed=339041;
 RA Kitten P.D., Burholo P., Sakurai Y., Yamada Y.;
 RT "Structure of the amino-terminal portion of the mouse alpha 1(IV)
 RT collagen chain and the corresponding region of the gene.";
 RL J. Biol. Chem. 263:8706-8709(1988).
 CC 1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A CHECKERBOARD
 CC NETWORK TOGETHER WITH LAMININS, PROTEGLYCANNS AND INACTIN/
 CC NIDOGEN.
 CC 1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOPFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC 1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS. FREQUENT INTERROPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC 1- PIM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC 1- PIM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BINDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.slb.ch/announcements>
 CC or send an email to license@isb.slb.ch).
 CC -----

DR EMBL: J03758; AAA57449.1;
 DR EMBL: M23333; AAA51625.1;
 DR EMBL: J04694; AAA50292.1;
 DR EMBL: X06777; CAA29946.1;
 DR EMBL: X02201; CAA26192.1;
 DR EMBL: M15812; AAA37340.1;
 DR EMBL: M14042; AAA37342.1;
 DR EMBL: M12879; AAA37343.1;
 DR EMBL: M13024; NOT_ANNOTATED_7DUS.
 DR EMBL: M13025; NOT_ANNOTATED_7DUS.
 DR EMBL: M13026; AAA37344.1;
 DR EMBL: M13027; AAA37345.1;
 DR EMBL: M13043; AAA37346.1;
 DR EMBL: J04448; AAA37437.1;
 DR EMBL: A33525; GCM548.
 DR MGI: 88454; Col4a1.
 DR InterPro: IP000087; Collagen.
 DR InterPro: IP001442; Procollagen*4.
 DR Pfam: PF01491; Collagen*4.
 DR Pfam: PF01413; C4_2.
 DR Pfam: PF000007; Collagen*4.
 DR Pfam: PF003923; Procollagen*4.
 DR Pfam: PF003923; Procollagen*4.
 DR SMART: SM00111; C4_2.
 DR SMART: SM00111; C4_2.
 KW Repeat: Hydroxylation; Glycoprotein; Collagen; Signal.
 KW Repeat: Hydroxylation; Glycoprotein; Collagen; Signal.
 FT PROPEP 1 28
 FT CHAIN 173 1669 AMINO TERMINAL PROPEPTIDE (7S DOMAIN).
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
 FT DISULFID 1460 1561 OR 1548 (BY SIMILARITY).
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
 FT DISULFID 1505 1511 BY SIMILARITY.

KI human tissues.
 KI J. Biol. Chem. 269:23013-23017(1994).
 KI [2]
 KI REVISIONS.
 KA Leinonen A.;
 KI Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 KI [4]
 KI SEQUENCE FROM N.A.; VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;
 KI O-1215; S-1277; T-1310; E-1334; E-1347 AND O-1661. AND VARIANTS R-433;
 KI E-162; Y-326; H-408; E-451; L-574; E-1269 AND P-1474.
 KI MEDLINE-21064696; PubMed 11134255;
 KA Heidet L., Arrondel C., Forcster L., Cohen-Solal L., Muller G.,
 KA Gauthier B., Stavrou C., Gullier M.C., Auriant C.;
 KI "Structure of the human type IV collagen gene COL4A4 and mutations in
 KI autosomal Alport syndrome."
 KI J. Am. Soc. Nephrol. 12:97 106(2001).
 KI [4]
 KI SEQUENCE OF 1454-1670 FROM N.A. AND PARTIAL SEQUENCE.
 KI MEDLINE-93015826; PubMed-1400291;
 KA Guinness S., Bernal D., Garcia-Soto M., Elena S.F., Saus J.;
 KI "Exon/intron structure of the human alpha 3(IV) gene encompassing the
 KI Goodpasture antigen (alpha 3(IV)NC1). Identification of a potential alpha
 KI antigenic region at the triple helix/NC1 domain junction."
 KI J. Biol. Chem. 267:19780-19784(1992).
 KI [6]
 KI SEQUENCE OF 1454-1670 FROM N.A.
 KI MEDLINE 91354570; PubMed-1882840;
 KA Morrison K.E., Matsuyama M., Yang-Feng J.L., Roeder S.L.;
 KI "Sequence and localization of a partial cDNA encoding the human alpha
 KI 3 chain of type IV collagen."
 KI Am. J. Hum. Genet. 49:545-554(1991).
 KI [6]
 KI SEQUENCE OF 1331-1670 FROM N.A.
 KI TISSUE-Kidney;
 KI MEDLINE-92147878; PubMed 1737849;
 KA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.;
 KA Pusey C.D.;
 KI "Molecular cloning of the human Goodpasture antigen demonstrates it
 KI to be the alpha 3 chain of type IV collagen."
 KI J. Clin. Invest. 89:592-601(1992).
 KI [7]
 KI SEQUENCE OF 1644-1670 FROM N.A.
 KI TISSUE-Kidney;
 KI Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
 KI [8]
 KI SEQUENCE OF 1449-1670, AND ALTERNATIVE SPLICING.
 KI TISSUE-Kidney;
 KI MEDLINE-94124547; PubMed 8294492;
 KA Peng L., Xia Y., Wilson C.H.;
 KI "Alternative splicing of the NC1 domain of the human alpha 3(IV)
 KI collagen gene: differential expression of mRNA transcripts that
 KI predict three protein variants with distinct carboxyl regions."
 KI J. Biol. Chem. 269:2342-2348(1994).
 KI [9]
 KI SEQUENCE OF 1-29 FROM N.A.
 KI MEDLINE-98196854; PubMed 9537506;
 KA Momota R., Sugimoto M., Ohashi T., Kitasawa K., Yoshida H.;
 KA Minomura Y.;
 KI "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
 KI alpha4(IV) collagen chains are arranged head-to-head on chromosome
 KI 2q36."
 KI FEBS Lett. 424:11-16(1998).
 KI [10]
 KI ALTERNATIVE SPLICING.
 KI MEDLINE-94280184; PubMed-8505332;
 KA Bernal D., Guinness S., Saus J.;
 KI "The human mRNA encoding the Goodpasture antigen is alternatively
 KI spliced."
 KI J. Biol. Chem. 268:12090-12094(1993).
 KI [11]
 KI VARIANT PRO-1474.
 KI MEDLINE-95078827; PubMed 7987301;

RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
 RA Bartelds A., Monnens L.A.H., van Oost B.A., Reumer H.G.;
 KA Roeder S.T., Smeets H.J.M.;
 KI "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
 KI recessive Alport syndrome."
 KI Hum. Mol. Genet. 3:1269-1274(1994).
 KI [1]
 KI FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 KI GLomerular BASEMENT MEMBRANES (GBM), FORMING A "CHICKEN-WIRE"
 KI MESHWORK TOGETHER WITH LAMININS, PROTEOLYANS AND ENACTIN/
 KI NIDDSIN.
 CC [1]
 CC SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC [1]
 CC SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC [1]
 CC ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: 1 (SHOWN HERE), 2/5 AND
 CC 3/15, ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
 CC C-TERMINAL NC1 DOMAINS.
 CC [1]
 CC TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLocalIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
 CC COCHLEA, LUNG AND BRAIN.
 CC [1]
 CC DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NON-COLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERACTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC [1]
 CC PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC [1]
 CC PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL
 CC PTM-TYPE IV COLLAGEN SITE.
 CC [1]
 CC ARE INVOLVED IN INTER- AND INTRACELLULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC [1]
 CC PTM: PHOSPHORYLATED BY THE Goodpasture antigen-binding protein.
 CC [1]
 CC DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
 CC [1]
 CC DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE 1
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
 CC GLomerulonephropathy CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
 CC MALES AND FEMALES.
 CC [1]
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announc/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: X60011; CAA56335.1;
 DR EMBL: A228487; CAC36101.1; JOINED.
 DR EMBL: A228488; CAC36101.1; JOINED.
 DR EMBL: A228489; CAC36101.1; JOINED.
 DR EMBL: A228490; CAC36101.1; JOINED.
 DR EMBL: A228491; CAC36101.1; JOINED.
 DR EMBL: A228492; CAC36101.1; JOINED.
 DR EMBL: A228493; CAC36101.1; JOINED.
 DR EMBL: A228494; CAC36101.1; JOINED.
 DR EMBL: A228495; CAC36101.1; JOINED.
 DR EMBL: A228496; CAC36101.1; JOINED.
 DR EMBL: A228497; CAC36101.1; JOINED.
 DR EMBL: A228498; CAC36101.1; JOINED.
 DR EMBL: A228499; CAC36101.1; JOINED.
 DR EMBL: A228500; CAC36101.1; JOINED.
 DR EMBL: A228501; CAC36101.1; JOINED.
 DR EMBL: A228502; CAC36101.1; JOINED.
 DR EMBL: A228503; CAC36101.1; JOINED.
 DR EMBL: A228504; CAC36101.1; JOINED.
 DR EMBL: A228505; CAC36101.1; JOINED.
 DR EMBL: A228506; CAC36101.1; JOINED.
 DR EMBL: A228507; CAC36101.1; JOINED.

[illegible][illegible]

[illegible][illegible]

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	F26	F27	F28	F29	F30	F31	F32	F33	F34	F35	F36	F37	F38	F39	F40	F41	F42	F43	F44	F45	F46	F47	F48	F49	F50	F51	F52	F53	F54	F55	F56	F57	F58	F59	F60	F61	F62	F63	F64	F65	F66	F67	F68	F69	F70	F71	F72	F73	F74	F75	F76	F77	F78	F79	F80	F81	F82	F83	F84	F85	F86	F87	F88	F89	F90	F91	F92	F93	F94	F95	F96	F97	F98	F99	F100	F101	F102	F103	F104	F105	F106	F107	F108	F109	F110	F111	F112	F113	F114	F115	F116	F117	F118	F119	F120	F121	F122	F123	F124	F125	F126	F127	F128	F129	F130	F131	F132	F133	F134	F135	F136	F137	F138	F139	F140	F141	F142	F143	F144	F145	F146	F147	F148	F149	F150	F151	F152	F153	F154	F155	F156	F157	F158	F159	F160	F161	F162	F163	F164	F165	F166	F167	F168	F169	F170	F171	F172	F173	F174	F175	F176	F177	F178	F179	F180	F181	F182	F183	F184	F185	F186	F187	F188	F189	F190	F191	F192	F193	F194	F195	F196	F197	F198	F199	F200	F201	F202	F203	F204	F205	F206	F207	F208	F209	F210	F211	F212	F213	F214	F215	F216	F217	F218	F219	F220	F221	F222	F223	F224	F225	F226	F227	F228	F229	F230	F231	F232	F233	F234	F235	F236	F237	F238	F239	F240	F241	F242	F243	F244	F245	F246	F247	F248	F249	F250	F251	F252	F253	F254	F255	F256	F257	F258	F259	F260	F261	F262	F263	F264	F265	F266	F267	F268	F269	F270	F271	F272	F273	F274	F275	F276	F277	F278	F279	F280	F281	F282	F283	F284	F285	F286	F287	F288	F289	F290	F291	F292	F293	F294	F295	F296	F297	F298	F299	F300	F301	F302	F303	F304	F305	F306	F307	F308	F309	F310	F311	F312	F313	F314	F315	F316	F317	F318	F319	F320	F321	F322	F323	F324	F325	F326	F327	F328	F329	F330	F331	F332	F333	F334	F335	F336	F337	F338	F339	F340	F341	F342	F343	F344	F345	F346	F347	F348	F349	F350	F351	F352	F353	F354	F355	F356	F357	F358	F359	F360	F361	F362	F363	F364	F365	F366	F367	F368	F369	F370	F371	F372	F373	F374	F375	F376	F377	F378	F379	F380	F381	F382	F383	F384	F385	F386	F387	F388	F389	F390	F391	F392	F393	F394	F395	F396	F397	F398	F399	F400	F401	F402	F403	F404	F405	F406	F407	F408	F409	F410	F411	F412	F413	F414	F415	F416	F417	F418	F419	F420	F421	F422	F423	F424	F425	F426	F427	F428	F429	F430	F431	F432	F433	F434	F435	F436	F437	F438	F439	F440	F441	F442	F443	F444	F445	F446	F447	F448	F449	F450	F451	F452	F453	F454	F455	F456	F457	F458	F459	F460	F461	F462	F463	F464	F465	F466	F467	F468	F469	F470	F471	F472	F473	F474	F475	F476	F477	F478	F479	F480	F481	F482	F483	F484	F485	F486	F487	F488	F489	F490	F491	F492	F493	F494	F495	F496	F497	F498	F499	F500	F501	F502	F503	F504	F505	F506	F507	F508	F509	F510	F511	F512	F513	F514	F515	F516	F517	F518	F519	F520	F521	F522	F523	F52
--	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----

[illegible]

```

Db 640 K-LVPLPGPPAR-GLKSPK-----FDGPGDKPTFTVRNGI-PQKCAVAGDGLFT 692
QY 594 DDV-FKIPKRNPPDQVITSGPAGGEPINPOL-SEEMQDQAFKQLIILPMAMIC 651
Db 693 GPPKRVKVG-LPDQMGFCTPGR---PQFNGI-PNNVAGQKQEGVAILKQLI 746
QY 652 LRPSNNRM-IGSGRHEFRNNNPLF--PRIPVE---QF--LSNRD-FKRIY 698
Db 747 L-PG-----PQKGSIGVGVGNGALGIPGLQGFCEGPFETLPSVG 794
QY 699 -----PQMGF-GRELEFGVPSNMKGVNIVNMK-----SNSMIFK 747
Db 795 SPGVGIGTPNARGPQPGQPTLSPGIGKEGFTPTGLMKGKQKQKQKQKQKQKQK 854
QY 738 REAGQPEEMKLRFGSDMLPAQKQVPLFPGEPQGVGMGPPLPNSGHSNGL 797
Db 855 GQSDI-----PG-----EPGQDAGGIPGFPKSGEMVMSHPQSPGSGAGL 901
QY 798 KNLKRPICPPKRNKSHMPPPLINSSNTSLNATPVQKGLKRLISVATSVHS 857
Db 902 -----PQKGD---HQPSSSGPGD-----GLKQKQKQKQKQKQKQK 934
QY 858 PG-----INLEKFTMLQVPSMLG---SPSNLKSPQTESLA---GMALPAAASIKS 907
Db 945 PQSMKQVMSMKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 994
QY 908 PVLVSAASAVHLKSPSLAP--SPGNTSSKPPQS-GLIPENKAPILMASPMIG 964
Db 995 PQLSLPQCA-----PQLPAPKQSVKQMLDGLFEGEGVIGPQVQSPPLPQKQK 1047
QY 964 NVESGPP-----PPINQVAVNINQSNISLFTYIMFFETL 1001
Db 1048 EKGAGGPGTIGIPQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1096
QY 1002 SONPLISMMSKPMSPSTPIYHDAIKTVAASIDSPAP-SPNLSNMKMGMINI 1060
Db 1099 KQSPQSV-----QYQSGSL-----PQKQKQKQKQKQKQKQKQKQKQKQK 1145
QY 1061 QNPLSGNPVYPLPSMKTQPLSHSNQMP--PNAVQ---QNIPIGVIMQPL 1114
Db 1146 GLPFTGP-----TGPQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1181
QY 1114 MSHNIMH-----GSEPPMPV--QGNMGF-----PQFPPVQSPQGVPPHNPSSQ 1162
Db 1182 KQDKQSKGEGVPPGLAGSPGIPQSGEGEMGPPQDQDPLPSGHA---DQKQKQK 1248
QY 1163 GSPFGMGFP--FGPIGPPSNLPQSSAAALCKKQPGCGPSTFVLGSMSTVDHILQ 1221
Db 1239 GP-QGQFGLPPLPQPMG-VPGIP--GLGVGKQKQKQKQKQKQKQKQKQKQK 1289
QY 1222 EVIRGATGIEFDLSRTIPSEKPSQTLQVPPKQVPPKQKQKQKQKQKQKQKQ 1281
Db 1290 GM-FGIGSSG-----TQSKQMG-----PQV-VPGQPGKPI-PQLGFTKQKQK 1345
QY 1282 PRMGLIPKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 1341
Db 1346 VPSAKPLP--GPPGPPHYDILKQ-FEGLFG--PEQPPGL--KQIGQ 1376
QY 1342 PQGQTLMSNPAVGMIPKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 1399
Db 1377 PG-----PQSGVIGLVGIPGPIPGQDQAPQKQKQKQKQKQKQKQKQKQK 1415
QY 1400 MRPKMAIWMGSPSGQGNIG 1422
Db 1416 TQPKRPQPPPKPIPKSMMPG 1448

```

```

DT 01-FEB-1996 (Rel. 33, last sequence update)

```

15-JUN-2002 (Rel. 41, last annotation update)
 DE Collagen alpha 1(VII) chain precursor (long-chain collagen) (1)
 DE Collagen.
 ON COL7A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE 94327588; PubMed 8051117;
 RT "Cloning of human type VII collagen, complete primary sequence of the
 alpha 1(VII) chain and identification of intrachain polymorphisms.";
 RL J. Biol. Chem. 269:20256-20262(1994).
 RN [2]
 RP SEQUENCE OF 128-1494 FROM N.A., AND PARTIAL SEQUENCE.
 RA MEDLINE 93338477; PubMed 1407247;
 RA Cristiano A.M., Rosenbaum J.M., Chung Hwee L.C., Parente M.A.,
 RA Wootley D.L., Pan T.C., Zhang K.Z., Chu M.-L., Furusoe K.E.,
 RA Uitto J.;
 RT "The large non-collagenous domain (NC-1) of type VII collagen is
 RT amino-terminal and chimeric. Homology to cartilage matrix protein,
 RT the type III domain of fibronectin and the A domains of von
 RT Willebrand factor.";
 RL Hum. Mol. Genet. 1:475-481(1992).
 RN [3]
 RP SEQUENCE OF 815-1415 FROM N.A.
 RA MEDLINE 91334380; PubMed 1871109;
 RA Parente M.G., Chung L.C., Kyriacou J., Wootley D.L., Wynn K.M.,
 RA Bauer E.A., Matisoff M.G., Chu M.-L., Uitto J.;
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the
 RT gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6941-6945(1991).
 RN [4]
 RP SEQUENCE OF 469-1255 FROM N.A.
 RA MEDLINE 9107742; PubMed 1469284;
 RA Gammou W.R., Abernethy M.L., Padilla K.M., Pitsayach P.S.,
 RA Cook M.E., Wright J., Briggman R.A., Hunt S.W. III;
 RT "Noncollagenous (NC1) domain of collagen VII resembles and interacts
 RT with extracellular matrix.";
 RL J. Invest. Dermatol. 99:691-696(1992).
 RN [5]
 RP SEQUENCE OF 340-675 FROM N.A.
 RA TISSUE=Retalimocytes;
 RA MEDLINE 92231902; PubMed 1567409;
 RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
 RT "Molecular cloning and characterization of type VII collagen cDNA.";
 RL Biochem. Biophys. Res. Commun. 183:958-964(1992).
 RN [6]
 RP SEQUENCE OF 205-2944 FROM N.A.
 RA MEDLINE 93271985; PubMed 8499916;
 RA Greenspan D.S.;
 RT "The carboxyl-terminal half of type VII collagen, including the non-
 RT collagenous NC 2 domain and intron/exon organization of the
 RT corresponding region of the COL7A1 gene.";
 RL Hum. Mol. Genet. 2:273-278(1993).
 RN [7]
 RP SEQUENCE OF 1-87 FROM N.A.
 RA TISSUE=Placenta;
 RA MEDLINE 94375010; PubMed 8088784;
 RA Cristiano A.M., Hoffman G.G., Chung Hwee L.C., Lee S., Chenu W.,
 RA Uitto J., Greenspan D.S.;
 RT "Structural organization of the human type VII collagen gene (COL7A1),
 RT composed of more exons than any previously characterized gene.";
 RL Genomics 21:169-179(1994).
 RN [8]
 RP REVIEW ON GENE VARIANTS.
 RA MEDLINE 98041696; PubMed 9375848;
 RA Jorvikallio A., Polkunen J., Uitto J.;
 RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in
 RT the type VII collagen gene (COL7A1)."


```

1b 1628 GPRGRGDEVCGRKDECPEDPDRHGLFKGRKCH-----KGL-RCAPGVKQ-----PVSRKRLQGS 1676
QY 298 LPPPGTGPSPNTPNNRAVFWVSQGNSSSAVKKAPPPVSSGCHPTTGLNPLNLSQPLD 357
1b 1677 -DVEGDEKRNSSP-----GSSNGKRGDRGEGEP-----GPRGRVLVLDLQD AHEKCE 1720
QY 358 HREHSGLFLK-----DIOHMLFPDEKFFTAQS--GGPQUNGV----- 394
1b 1721 FQHBGQGRPKGKDPDLPIAPABERGTEGRGDPGPGDGVKDPACERKDHGHTDGLR 1780
QY 495 -LDDP--GKKPRSPVQAMMAUSUSLGRK-----FGRRTVGAPEGRHGHVPESP 442
1b 1781 SGDLGRKGAAGPSNP-----MGAAKAGDTGRTGLPDRGEGDLP--GNSPEGLTAKP 1842
QY 443 DEMVPPSSNSQSTLGGHLLDHTPEQJLAMI-KLQDFEPEKRRKQEVVVVQVSTLQMMV 502
1b 1843 GEDKRPGLNCKNEPDP-----GEGKRGKGRGYSQASGRDQ--- 1869
QY 503 HOHTRGV-----VRGP-----PYQMTSEGMALPGTPEPS---GLNMRHS 543
1b 1870 -KDDPKGRGAPOLLAPGFTGLPGVGPQGCHV--VGGDGRKGGHRTGSKTSDG 1925
QY 544 LP-PRMAPPRNMDSQMLPFCAGMLNS-----EMRGNVNTASRGLSVSW 592
1b 1926 LPGRGRLRGEPQSVYVNDLLETAGKASALRELVETWDESSQPLVPERKRG- 1979
QY 593 PDGVPLTPDKRNPPG--QGTFSNGP-----RGR-FNDVGLSEEMVQOLATKQJLTPG 646
1b 1980 -PKDSSGEGPCKEKEGLTGRGEGRLKDDGDRGQD-----PRG 2018
QY 647 MAMEGTPSSNMENMLTPG--SQRMEPGNNDLPRIIP-----VEGPI 666
1b 2019 LALGERGP-----PQSLAFEPG-KPGITGLDFRAGVGPAGPDEKGRGKGR 2069
QY 687 SPSPGQEPKTPQMGHRELFEGWPSQKQKQVNLVNNSSQMLPKMPLAGAPPE 746
1b 2070 GEGRCRGPPLPGTFRGPG-----PGRKVSVD--EPGRLSGEGPPLKKA----- 2115
QY 747 MLKRGGS--DMLPAQCKVPLPFGHTFOVEYCMGPEPFLPMSGGRPSNSGLNLRPTG 805
1b 2116 -KGEPSNCDQGRPKGRGVYGLKDDHGER-----GPR-----QGGNGLDTPKRMAS 2162
QY 806 PDORTNRLSHPEPLPLNSSNPTSLNTAIPVQKGLKRLDLSVANSVSHATNPLKS 865
1b 2163 PEKPKPGLGPRGPGVGGHGP-----GPR----- 2188
QY 866 PTHHGVSPMLGSPSGLKSPQPSQLAG-----LAPRAAASLKSIP---VF 911
1b 2189 -GAPGLAPAG-----PQSPSLKDEPDETPGPRGRLTSPGLAVMLD-GRSRTS-IV 2238
QY 912 GSAASIVHLKSPSLPAPSPHWSKPPVQ-----SPGTPPNKATLMASSIA 960
1b 2239 GPVG-----STGTPG--QVGETGPKAPGRKQDASIKDKGSPGV-----GSA 2281
QY 961 MLCNVSSGCHPPT-ASQFASVNPQS-----LPSSTPVMPPEPLSONHSTMSKMS 1014
1b 2282 LPQVVGKGGEPGLGAGVAVGLPEAKGCKGAPGLAGLWGEPR---GAK 2340
QY 1015 KFMASSTPLVHATKTVASSDDDSPPAR--SEPLPSMNMMPGNGTONTQKSGNPPV 1073
1b 2331 DQGLPDRGCRGAGRAGEHGDGELGOKGAPDKPKGKQGVV---PSTGDP--- 2381
QY 1074 MPTLSMGMLTQPLSHSNOMSPNAVGNIPLPHGVEMPGJLMNSNPIMCHSS-FTTVV 1133
1b 2382 -PGRVVGKDDLDELPLPAPG-----VGGP-LQIT--PRG 2413
QY 1134 RKGFPQGPVVSPPQGVPTPHNPSQOG--SPPGMGFPGEPLATPSNIPOVSAIAA 1191
1b 2414 EMQGP-----GPRSGRLAGPRGRGIP--GLQDPG----- 2443
QY 1192 LCAVGGGGRNSTVLGNSKMPSTFTPDQLQEVLRPAGATIPEDTSLTPSKTSVLY 1251
1b 2444 -PFGVGPVPGASGLKGRK-----GPR-----GVGHPG----- 2459

```

```

QY 1252 FPRGE-----VPRKQPGG--PG----- 1267
1b 2470 -PRGPRGCTLRGDEDRPQDEPGRGHTGPPSGKGRGKGVASAR-KGKRLNSAVLIR 2528
QY 1268 TQSHHGMMGRCVAPRMGLALPG-----MAGGVVLDLP--GLVAPSRHNPW 1318
1b 2529 PRGRGKQJMEGRGP-----GLDQKGRGPDNDPDKSKRGEGKSA-GLAG--- 2578
QY 1319 PAVLQGMGSRHHRMSDAGSLMPSGPIILMSNA----- 1352
1b 2579 -LMDLGLPQGV---PAAAGLWGP---GSPKGVVGLGKRLGVGPRHGLKGR 2628
QY 1353 -AAVGMILNG-DRGA-----GLYHGVGSGPM---MMSDGMHPPQNMILP 1397
1b 2629 GVGAGGGLDEPKRDKGRFAGPRPRLAKRGKMGEGGVVQVSTAVYKRLTQKTRGPD 2688
QY 1398 PGMPPROMAD-----VGMGRSGGPGNPN 1423
1b 2689 GQPRKQDQKGRGRCFGLGGRPSNGS 2720

```

RESULT 10

CA24-CAEEL

ID CA24-CAEEL STANDARD: PRI: 1758 AA.

AC P17140: Q19098; Q19099;

DT 01-AUG-1990 (rel. 15, created)

DT 01-OCT-1994 (rel. 30, last sequence update)

DT 15-JUN-2002 (rel. 41, last annotation update)

DE collagen alpha 2(IV) chain precursor (lethal protein 2).

GN IEF-2 OR CUB-1 OR FOLG12.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromodorea; Kladistida; Kladistoida.

OC Rhabditida; Peridornae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A., AND FUNCTION.

RC STRAIN=Bristol N2;

RX MEDLINE=94012964; PubMed 7691828;

KA Stibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;

KT Genetic identification, sequence, and alternative splicing of the

RL Caenorhabditis elegans alpha 2(IV) collagen gene.*

RL J. Cell Biol. 123:255-264(1993).

RN [2]

RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=9008929; PubMed 2794871;

KA Guo X., Kramer J.M.;

KT *The two Caenorhabditis elegans basement membrane (type IV) collagen

RT genes are located on separate chromosomes.*

RL J. Biol. Chem. 264:17574-17582(1989).

RN [3]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING

RC STRAIN=Bristol N2;

RA Wu X., Le T.T.;

RB Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.

RN [4]

RP VARIANTS.

RX MEDLINE=94120591; PubMed 8045258;

KA Stibley M.H., Graham P.L., von Mendel N., Kramer J.M.;

RT Mutations in the alpha 2(IV) basement membrane collagen gene of

RT Caenorhabditis elegans produce phenotypes of different severities.*

RL EMBO J. 13:3278-3285(1994).

CC -1- FUNCTION: Collagen type IV is specific for basement membranes.

CC -1- VITAL for embryonic development.

CC -1- SOUNDNIT: FRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.

CC TYPE IV COLLAGEN FORMS A MESH LIKE NETWORK LINKED THROUGH

CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1

CC DOMAINS.

CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, 1/a (shown here) and 1/b are

CC produced by alternative splicing.

CC -1- DEVELOPMENTAL STAGE: Isoform 1 is predominant in embryos and

CC isoform 11 is predominant in the larvae and adults.


```

07 1248 TIOYFPRGIVPQKROUQUPR--GEFSIMQ NMMEDEADRMALAFQ-MQVGVTVTELEL 1405
08      | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
09 1636 -----ATG--AGAGAPGPAPAKNKNDIKSTGAAGTACGTGGTAATACATAC TGCTGG 1784
10
11 1406 TA-----PSMGMIPMPHVAITQSGMMQDPINRKMSAUSTMPASPTLMSSMAAVNM 1457
12      | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
13 1084 DKRTLEGEDLDKRGKGRDCE-----SLGLSP-----PEEPFAVEH GSHNSANS 1125
14
15 1458 ITRKRGRPA-----CLTYRQIVGSIGMMMSMQ NMMLTGUNIMLPYQMKRMAAL 1408
16      | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
17 1126 -PAGRPGRPASACAAKCKRIKIDFGSELCDFG-----FRCHTGF 1162
18
19 1499 VCGMGSESGCNGNG 1422
20      | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
21 1164 VGCPVPFGPGHPFG 1176
22
23 RESULT 14
24 CAZB_MOUSE
25 ID CAZB_MOUSE STANDARD PRI: 1650 AA
26 AC Q64739; Q9Z1W0;
27 DT 15-JUL-1998 (rel. 36, Created)
28 DT 10-MAY-2000 (rel. 39, Last sequence update)
29 DT 15-JUN-2002 (rel. 41, Last annotation update)
30 DE Collagen alpha 2(XI) chain precursor.
31 GN Coll1a2.
32 OS Mus musculus (Mouse).
33 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
34 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
35 OX NCBI_Taxid=10090;
36 RN (1)
37 RP SEQUENCE FROM N.A.
38 KC STRAIN:129/SvEv;
39 KA Rowen L., Qin S., Madan A., Loretz C., James R., Boas M., Max L.,
40 KA Hall J., Lasky S., Hood L.;
41 RT "Sequence of the mouse major histocompatibility locus class II
42 RT region.";
43 RL Submitted (OCT-1998) to the EMBL/GenBank/JDBJ databases.
44 RN [2]
45 RP SEQUENCE OF 1-1592 FROM N.A.
46 RC STRAIN:FVB/N, AND 129/SV; TISSUE:Cartilage;
47 RX MEDLINE=97135795; PubMed=8981932;
48 RA Vandenberg P., Vorlisto M.M., Ala-Kokko L., Prockop D.J.;
49 RT "The mouse coll1a2 gene, Some transcripts from the adjacent ext1a2a
50 RT gene extend into the coll1a2 gene.";
51 RI Matrix Biol. 15:359-367(1995);
52 CC
53 CC -1 FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILGENESIS BY
54 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY).
55 CC -1 SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
56 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST TRANSLATIONAL
57 CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) CAN ALSO BE FOUND INSIDE
58 CC OF ALPHA 3(XI)-(II) (BY SIMILARITY).
59 CC -1 ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY
60 CC ALTERNATIVE SPLICING.
61 CC -1 PTM: PRO-LINES AT THE THIRD POSITION OF THE REPEPTIDE REPEATING
62 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
63 CC -1 SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
64 CC -----
65 CC This SWISS-PROT entry is copyright . It is produced through a collaboration between
66 CC the European Bioinformatics Institute and the EMBL outstation at
67 CC the European Bioinformatics Institute. There are no restrictions on its
68 CC use by non-profit institutions as long as they control it in no way
69 CC modified and this statement is not removed, used by and for commercial
70 CC entities requires a license agreement (See http://www.isdb.ch/about/cv/
71 CC or send an email to jlicense@isb.sib.ch).
72 CC -----
73 DR EMBL: AF100956; AAC69905.1; -
74 DR EMBL: U16789; AAA67751.1; -
75 DR EMBL: U16790; AAA67752.1; -
76 DR MGDB: MG1:88447; Coll1a2.
77 DR InterPro: IPR000087; Collagen.
78 DR InterPro: IPR000886; Fib_collagen_C.
79 InterPro: IPR001791; laminin_G.

```


- OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Invertebrata;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae;
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-472 FROM N.A.
 RX MEDLINE=89025644; PubMed=3178744;
 RA Tromp G., Kivimäki H., Stacey A., Shkuta H., Balkwill F.L.,
 RA Jancsik R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the pro-alpha 1(I)
 RT chain of human type I procollagen.";
 RL Biochem. J. 253:919-922(1988).
 RN [2]
 RP SEQUENCE OF 1-181 FROM N.A.
 RX MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Moradot M.,
 RA Myers G., Williams G., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons.";
 RL Nucleic Acids Res. 12:137-140(1984).
 RN [3]
 RP SEQUENCE OF 162-401.
 RX MEDLINE=71038625; PubMed=5529814;
 RA Clark E.M., Bernstein P.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706(1970).
 RN [4]
 RP SEQUENCE OF 265-268.
 RX MEDLINE=71001508; PubMed=4419110;
 RA Morán P.H., Jacobs H.G., Stewart J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides from selected
 RT vertebrate collagens. A possible role of the carbohydrate in fibril
 RT formation.";
 RL J. Biol. Chem. 245:5042-5046(1970).
 RN [5]
 RP SEQUENCE OF 425-1464 FROM N.A.
 RX MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.G., Ramirez F., Eickbush F.P.,
 RA Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5221(1983).
 RN [6]
 RP SEQUENCE OF 1225-1454 FROM N.A.
 RX MEDLINE=88124208; PubMed=3440541;
 RA Meekelae J.K., Raustina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:3349-3349(1988).
 RN [7]
 RP SEQUENCE OF 1-84 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bernstein P., McKay J., Morishima J.K., Devaratnam S., Gethas R.E.;
 RT "Regulatory elements in the first intron contribute to
 RT transcriptional control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8859-8871(1987).
 RN [8]
 RP SEQUENCE OF 1-84 FROM N.A.
 RX MEDLINE=8510970; PubMed=2857713;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1 (I) cDNA gene,
 RT promoter structure, Alu repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2420(1985).
 RN [9]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033048; PubMed=2822714;
 RA Rossow C.M.S., Verrier W.P., da Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.J.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010958;
 RA Kivimäki H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrillar
 RT associated collagen (type IX), and network forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97253959; PubMed=9101290;
 RA Kivimäki H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrillar
 RT associated collagen (type IX), and network forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91374476; PubMed=1895412;
 RA Byers P.H., Wallis G.A., Walling M.C.;
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
 RL J. Med. Genet. 28:433-442(1991).
 RN [13]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97169389; PubMed=9016532;
 RA Delisle R.;
 RT "The human type I collagen mutation database.";
 RL Nucleic Acids Res. 25:181-187(1997).
 RN [14]
 RP VARIANT OF-11 CYS-1166.
 RX MEDLINE=86287390; PubMed=3016787;
 RA Cole D.H., Byers P.H., Steinmann M., Gellinas R.E.;
 RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
 RT change in one human pro alpha 1(I) collagen allele.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
 RN [15]
 RP VARIANT OF-11 ARG-569.
 RX MEDLINE=87222295; PubMed=3108247;
 RA Bateman J.F., Chan D., Walters L.D., Rogers L.O., Cole W.G.;
 RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
 RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
 RT collagen.";
 RL J. Biol. Chem. 262:7021-7027(1987).
 RN [16]
 RP VARIANT OF-11 CYS-926.
 RX MEDLINE=88033041; PubMed=3667599;
 RA Vogel R.E., Minor R.R., Etzold M., Prockop D.J.;
 RT "A point mutation in a type I procollagen gene converts glycine 746
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
 RT a lethal variant of osteogenesis imperfecta.";
 RL J. Biol. Chem. 262:14737-14744(1987).
 RN [17]
 RP VARIANT OF-11 ARG-842.
 RX MEDLINE=88248828; PubMed=3403560;
 RA Bateman J.F., Lammie S.R., Dahl H.H., Chan D., Cole W.G.;
 RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
 RT chain in lethal perinatal osteogenesis imperfecta: demonstration of
 RT the peptide defect by in vitro expression of the mutant cDNA.";
 RL J. Biol. Chem. 263:11627-11630(1988).
 RN [18]
 RP VARIANT OF CYS-1195.
 RX MEDLINE=89218628; PubMed=3244312;
 RA Liebhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister L.W.;
 RT "A cysteine for glycine substitution at position 1017 in an alpha
 RT 1(I) chain of type I collagen in a patient with mild dominantly
 RT inherited osteogenesis imperfecta.";
 RL Mol. Biol. Med. 5:197-207(1988).
 RN [19]
 RP VARIANT OF-11 VAL-416.
 RX MEDLINE=89254493; PubMed=2470760;
 RA Patterson F., Smiley E., Bonadio J.;

61 "RNA sequence analysis of a putative beta-actin gene from
 62 rat liver."
 63 J. Biol. Chem. 264:10084-10087(1989).
 64
 65 VARIANT of IV csp 1000.
 66
 67 METALINE 89400001, PubMed 27451202.
 68 Mettler J. C., Gagnon L. R., Gotsman G. S., Lewis M. B., Kozlowski A. J.
 69 "Polynucleotide phosphorylase from a type IV, isolation of a partial cDNA for
 70 gene alpha (II) isolated from a (GOLD) by RNA/RNA hybrid analysis."
 71 J. Biol. Chem. 264:11893-11896(1989).
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000


```

Db 1107 IKHROF-----SLOQP-----PDPPOSBOBPSGASGAGPPOHPSNGAPJOKU 1154
QY 1163 KGFAGLYIHNGPVGSPGMMSSMOGMMGPDOUNIMPIVOMPRGMADVMGMMAPSSGAPENP3 1422
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 1154 ---GIMHFGPGLGPPG-----PRGKGLAGVAFPGPQPPG 1187

```

Search completed: February 28, 2003, 16:05:41
 Job time : 42 secs

[illegible][illegible][illegible]

A. Receptor gene bankcodes: A44758; MIMID: 9427414; PDBID: 6B00044
A. Receptor sequence: amino acid (aa) coding sequence for the human serotonin-_{1A} receptor; The Genbank accession number: AF094494
J. Biol. Chem. 269: 12090-12094, 1994
A. Ligand: The human mRNA encoding the α -opioid receptor and its alternatively spliced isoforms; The Genbank accession number: AF5971; MIMID: 9380184; PDBID: 8F0542
A. Accession: AF5971

A. S. KATSON, *Interfacial acid sequences and structure*
A. Molinari, *Type II RNA*
A. Molinari, *14-3-3, V4.4, V4.5, V4.6*
A. Molinari, *Sequence evolution from M101 (M101:14-3-3)*
R. M. L. S. S. K. P. J. Molinari, M. J. Molinari, *14-3-3, V4.4, V4.5, V4.6*
A. Molinari, *Sequence and localization of a human alpha chain*

[illegible]

^a ϵ (30°C/mol): 1,10-diphenylbutane's synchronous and anticomparative responses observed upon an applied electric field were: $\Delta\epsilon_{\text{sync}} = 0.0013$ and $\Delta\epsilon_{\text{anti}} = 0.0001$ (1000 V/cm).

Keywords: alternative spinning; bamboo; nonwoven; cold bonding; cold melt; extrusion; extrusion spinning; extrusion nonwoven; structural component of extruded; at basement membrane; in kidney; support family; collagen alpha 1(IV) chain

E-1 28/04/2019: animal sequence #status produced /MA
 E-2 16/03/2019: cytokinin alpha (1V) chitin, lipid species form #status produced /MA
 E-3 19/04/2019: amino-terminal nonhydroxyl, NH₂-NH₂
 E-4 14/05/2019: intercalated, locked
 E-5 29/04/2019: cell attachment (R G) mod II
 E-6 06/06/2019: cell attachment (R G) mod II
 E-7 14/06/2019: cell attachment (R G) mod II

Run	Model	Model description	Model parameters	Model results
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10
11	11	11	11	11
12	12	12	12	12
13	13	13	13	13
14	14	14	14	14
15	15	15	15	15
16	16	16	16	16
17	17	17	17	17
18	18	18	18	18
19	19	19	19	19
20	20	20	20	20
21	21	21	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
25	25	25	25	25
26	26	26	26	26
27	27	27	27	27
28	28	28	28	28
29	29	29	29	29
30	30	30	30	30
31	31	31	31	31
32	32	32	32	32
33	33	33	33	33
34	34	34	34	34
35	35	35	35	35
36	36	36	36	36
37	37	37	37	37
38	38	38	38	38
39	39	39	39	39
40	40	40	40	40
41	41	41	41	41
42	42	42	42	42
43	43	43	43	43
44	44	44	44	44
45	45	45	45	45
46	46	46	46	46
47	47	47	47	47
48	48	48	48	48
49	49	49	49	49
50	50	50	50	50
51	51	51	51	51
52	52	52	52	52
53	53	53	53	53
54	54	54	54	54
55	55	55	55	55
56	56	56	56	56
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
61	61	61	61	61
62	62	62	62	62
63	63	63	63	63
64	64	64	64	64
65	65	65	65	65
66	66	66	66	66
67	67	67	67	67
68	68	68	68	68
69	69	69	69	69
70	70	70	70	70
71	71	71	71	71
72	72	72	72	72
73	73	73	73	73
74	74	74	74	74
75	75	75	75	75
76	76	76	76	76
77	77	77	77	77
78	78	78	78	78
79	79	79	79	79
80	80	80	80	80
81	81	81	81	81
82	82	82	82	82
83	83	83	83	83
84	84	84	84	84
85	85	85	85	85
86	86	86	86	86
87	87	87	87	8

[illegible]

[illegible][illegible]

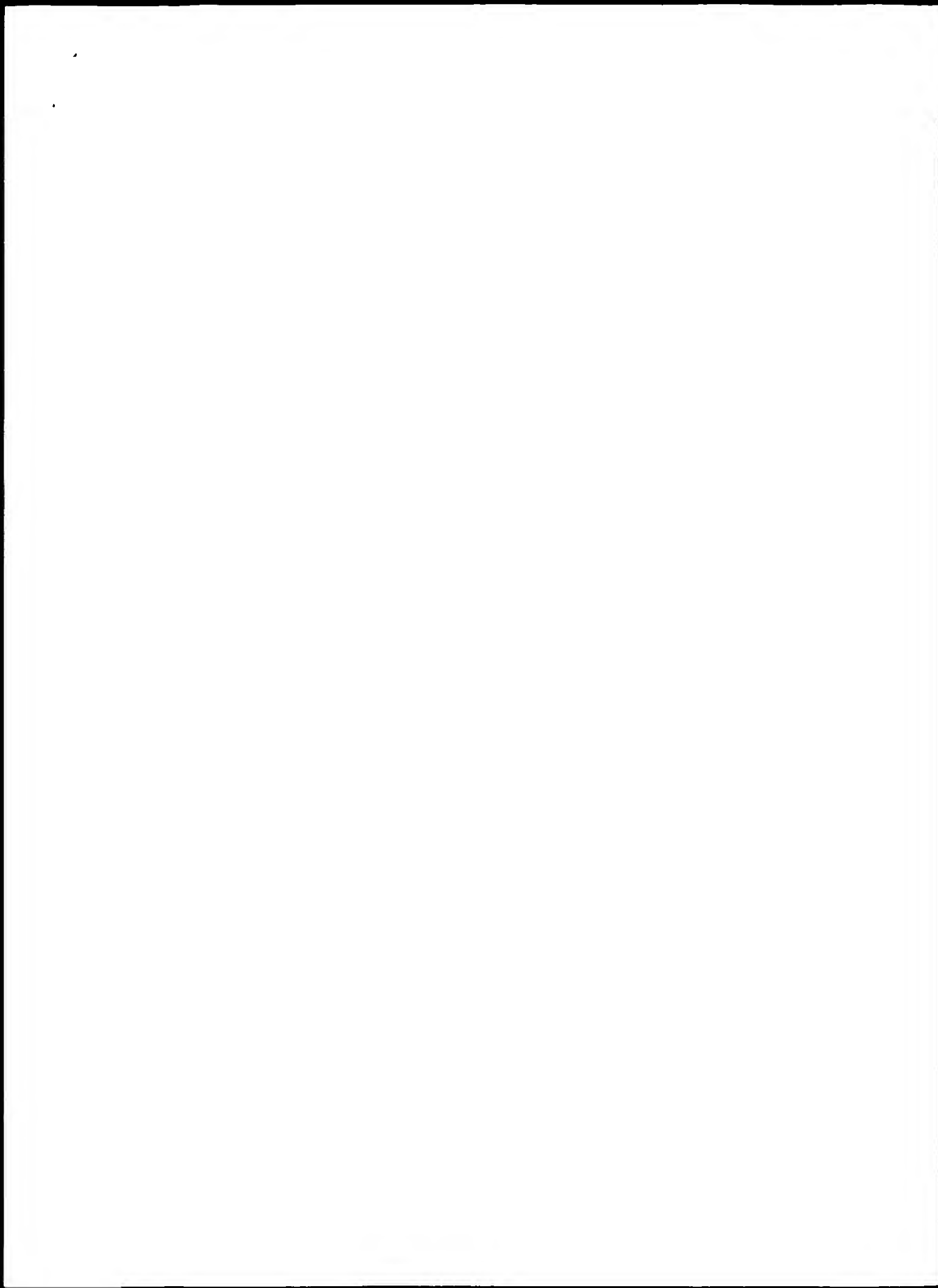
[illegible][illegible]

A: Reference number: 220611
 A: Accession: 129431
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: RNA
 A: Residues: 1-1759 (507X)
 A: Cross-references: EMBL:05442; PDB:AAA9215; 1: GSP08-030028; 1: S1-00112-5a
 A: Experimental source: strain Bristol N2; clone p11G12
 A: Genes:
 A: Gene: GSP01012-5a
 A: Map position: X
 A: Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 757/3
 A: Superfamily: collagen alpha 1(I) chain
 Query Match 6.78; Score 516.5; DB 2; Length 1759
 Best Local Similarity 23.84; Pred. No. 6; Lp 15
 Matches 356; Conservative 104; Mismatches 537; Indels 15; Gaps 82

QY 238 PPAVANDUNSSONTROPTPTTAVAP-----KIAAPPTPDREPTVFNKIP 267
 DB 98 PGPFGNNDGNI-----PDPFGPAPGMDGNGTICAGATPDRGDPNMP-PP 149
 QY 288 SV-----GSPASTPLDP-----EIGFENST--PNNRAVTPVSQ--SNSSAAFKAP 333
 DB 150 GMRKAKPPATGAAAGAGGKATGAGKMPGLPDSNKGTPGKGGGAGGAGAGAG 209
 QY 334 PVSSSLPPTLGGNDGLSGEOLHFRFSLQTRDQRMPLPDKERTAGSNTVQNG 394
 DB 210 EAGSPGNF-----GSPGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 229
 QY 394 VLD--GIGKKPRPQIAGMAAGSOSLG-----KG-PGPRIVAP-- 429
 DB 230 DTCAGMP--AATPGPLASIMSKGILGPKDIOEKKGKGGGMDGRTYFNNLSD 288
 QY 430 -----EPQCHREVPSPEPMVPSNNSGTLGPRLIMPTPLAW-KIQEELY 460
 DB 289 PAKKGGKGLSPKPKPKKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 542
 QY 481 EKKRKQDVVQGSLSQIMVHGHPRVGVKPPVQMTPSHWAPSLPTPLSLTMM 543
 DB 343 PGR-----IDPKG--NSHPGP-----PAG--TFNNAAI 569
 QY 541 PHLSP-----PRMAPIHNMG-----SOMRIPACMINSLMTP--NVPFASV 594
 DB 370 PPGDGPNGKPGPDGDPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 427
 QY 585 PGLSGSWPDVDFKIPDGRNP-----PQD-----GTFSPGR 637
 DB 428 PGADVVSQFSGITGLDQPKGPYKGTPOSIPGLPKPKPKPKGLDQAPKPKPN 487
 QY 618 -----GR-FTNPGLSEEMPOOLAKKULIPNNMAMGLK-- 654
 DB 488 GPEDSLNLPAPQGAGAPNGVYR-----DGVNTPGAGATKILGSDTNSAP 540
 QY 655 SAKMKKIPGSGHMHGPNP-LETRIP-----VAPLIS-----PSMGLPKEL 697
 DB 541 GIKGGKGLPYSQPGPGGGRLEGMGHWYDAG--DGLPQAGKPSDVPYDGL 599
 QY 698 PPGMG-----PRLERGMV-----PS-----AKGQVW 721
 DB 600 PDKGRTVLTILGPGPTGGLAKSEMPFGQGVNGLPAPSGVAPGAG--DQ 659
 QY 722 LNVNMSNSV-----MPPKRRFAPAPPEMLKLP-----GSSIMPLAKKVP 766
 DB 660 LPLSKRPQDGLGLGKMKAGVY-----GPDGRTWAKKAGLPLPGLPGLD 712
 QY 767 LPPGHPDQFYKGP--KPTLMSQCHNSNLTPLREPTG-- 712
 DB 713 MGGPAPKPVNAPPGPGLPGLPKGPGYKMGEGVQGPFGPLPKMKLSTLGP 772
 QY 818 PPLPLPSSNPTSLNAPVQKGLCKKPLDISAA--SOVHSGLNPKSTIMHGVQSL 875
 DB 773 PGLPGR-----GVPDQKGFQGVNGLPGLPKKPGVNDGLPGLM-- 819

QY 876 LASHSNLKSPQTSQAGMAGIAAASLSPVLSAAASVHAKSLSTLAS 943
 DB 820 VQVF--GQVSGTPGG-----LKIALPPTPTGAPKAPHP--GAPYKKAHPL 870
 QY 934 DSSKPTQ-----SPGIPNKKAPLMASVAMLVNSSTPPPLAS 976
 DB 871 SDPGYTGKGLAGLPVPRGSGSPVPGDGLP--GVKMGK--DGLPGLVETAK 925
 QY 977 ----QVAVSLPSSLSSTPYMTPELTLSQNTSLIMSSKSPVAGSSLPYHAKIV 1032
 DB 926 GILGAPGSGAPG-LPRAVY--PGRKNA-----GILV 957
 QY 1033 ASSIDP-----SLVAPSLPSSNNPNSGCTNIDNLSLNPVYVPL- 1077
 DB 958 PAKKGLDGLGLPGLNDFKPPVAVPMGTPDKKMGALPLGKGLASVPSKAPK 1017
 QY 1078 ----SPMGVPLSHSNQ--MP--SPNAVDPNLTPEVPMGGLMSNIIMGSSQPE 1127
 DB 1018 NGLPGRKGAPPAKQVQKGLAGLWQVPLKQV--GHSGLVAVCHKGLTGVYQV 1066
 QY 1128 LPPVQGRKPT-----GEPVVSPTVQ--VP--PPLNGLSG--GQSTVQMG 1173
 DB 1077 P--GKGLDGLPQAKGQAP--GSPQDGLPQPKGSGSYGQGLPDKMLP 1130
 QY 1174 ----EPGLGRSNIPOSSAALAGKPGGMDPSTVLEN--SMDSVPLPGL 1221
 DB 1131 POKKGLDQSGQGLSGAFLDQDPVYSG-----LPRKQAGLTPGQDKRMD 1184
 QY 1222 --FVLRGATGIPEDLSRTIPSEKPSQTLQVFP--KGF-----VPSKQV 1265
 DB 1185 KQENVEFGGLG-----LQDGRKFPAGLPVPTGKQSGVYQGLGLGSLKLS 1238
 QY 1266 PG--PQSHMGQMGQDAPKMGALAPMG--PGVATLPLATVAPSMGNN 1314
 DB 1239 PGPVQGLPGLSGK--MG--GLPMPQPGQSLAGVGPFG--ADVQGRK 1290
 QY 1317 RPPALQGGKMSPHHHHMSIAGSIMPGLTMSNPAVAGMTPKRGCPNLTHTG 1376
 DB 1291 LPQGRKESGLSHLPKAVGLKSGSGMPFPVAKGLATA--NGLPK--PGL 1347
 QY 1377 SPKMMMSQSMQVQNLIMLPQMRPQMAAVQMGQSPQVKNVNNMP 1426
 DB 1348 QVQ--LTLKGFVQAGAGLPQGPPTGLTGLKGRGLGLGHVAG--PAP 1393

Search completed: February 28, 2003, 16:06, 40
 Job time: 58 secs



Sequence version 5.1.13
Copyright (c) 1993-2003 CompuLink Ltd.

us-09-915-543-15

Run on: February 28, 2003, 16:06:34 : Search time: 22 seconds
(without alignment)

Database: US-09-915-543-15

Sequence: 1 MESSINKYSSPSMPPSP...ALVMDKSLINIMM 140

Scoring Table:

Match 10.0, Mismatch 0.5

Scored: 174566 scores, 4771826 residues

Total number of hits satisfying chosen parameters: 174566

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match: 08

Maximum Match: 1008

Listing first 45 summaries

Database:	Post-processed:	Score:	Match:	Length:	DB:	LF:	Page:	File:
1	us-09-915-543-15	507	6.6	2005	10	US-09-745-467B-3	1	us-09-745-467B-3
2	us-09-915-543-15	507	6.6	2005	10	US-09-745-467B-3	1	us-09-745-467B-3
3	us-09-915-543-15	492	6.4	1806	10	US-09-745-467B-3	1	us-09-745-467B-3
4	us-09-915-543-15	481	6.0	1567	10	US-09-745-467B-3	1	us-09-745-467B-3
5	us-09-915-543-15	461	4.7	2507	9	US-09-819-106A-2	1	us-09-819-106A-2
6	us-09-915-543-15	456	4.6	4019	9	US-09-854-134-425	1	us-09-854-134-425
7	us-09-915-543-15	456	4.6	4019	9	US-09-854-134-425	1	us-09-854-134-425
8	us-09-915-543-15	449	4.5	2402	9	US-09-745-467B-3	1	us-09-745-467B-3
9	us-09-915-543-15	445	4.5	1255	9	US-09-819-106A-2	1	us-09-819-106A-2
10	us-09-915-543-15	445	4.5	1255	9	US-09-819-106A-2	1	us-09-819-106A-2
11	us-09-915-543-15	427	4.4	2412	12	US-10-124-557-5	1	us-10-124-557-5
12	us-09-915-543-15	427	4.4	2412	12	US-10-124-557-5	1	us-10-124-557-5
13	us-09-915-543-15	416	4.1	2665	10	US-09-864-761-3124B	1	us-09-864-761-3124B
14	us-09-915-543-15	413	4.1	2441	12	US-10-124-557-5	1	us-10-124-557-5
15	us-09-915-543-15	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
16	us-09-915-543-15	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
17	us-09-915-543-15	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
18	us-09-915-543-15	410	4.0	1414	12	US-10-124-557-5	1	us-10-124-557-5
19	us-09-915-543-15	410	4.0	1414	12	US-10-124-557-5	1	us-10-124-557-5

Prod. No. is the number of results predicted by chosen parameters, have a score greater than or equal to the score of the result being predicted, and is derived by analysis of the total score distribution.

SUMMARIES

Prod. No.	Score	Match	Length	DB	LF	Page	File
1	507	6.6	2005	10	US-09-745-467B-3	1	us-09-745-467B-3
2	507	6.6	2005	10	US-09-745-467B-3	1	us-09-745-467B-3
3	492	6.4	1806	10	US-09-745-467B-3	1	us-09-745-467B-3
4	481	6.0	1567	10	US-09-745-467B-3	1	us-09-745-467B-3
5	461	4.7	2507	9	US-09-819-106A-2	1	us-09-819-106A-2
6	456	4.6	4019	9	US-09-854-134-425	1	us-09-854-134-425
7	456	4.6	4019	9	US-09-854-134-425	1	us-09-854-134-425
8	449	4.5	2402	9	US-09-745-467B-3	1	us-09-745-467B-3
9	445	4.5	1255	9	US-09-819-106A-2	1	us-09-819-106A-2
10	445	4.5	1255	9	US-09-819-106A-2	1	us-09-819-106A-2
11	427	4.4	2412	12	US-10-124-557-5	1	us-10-124-557-5
12	427	4.4	2412	12	US-10-124-557-5	1	us-10-124-557-5
13	416	4.1	2665	10	US-09-864-761-3124B	1	us-09-864-761-3124B
14	413	4.1	2441	12	US-10-124-557-5	1	us-10-124-557-5
15	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
16	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
17	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
18	410	4.0	1414	12	US-10-124-557-5	1	us-10-124-557-5
19	410	4.0	1414	12	US-10-124-557-5	1	us-10-124-557-5

ALIGNMENTS

Prod. No.	Score	Match	Length	DB	LF	Page	File
1	507	6.6	2005	10	US-09-745-467B-3	1	us-09-745-467B-3
2	507	6.6	2005	10	US-09-745-467B-3	1	us-09-745-467B-3
3	492	6.4	1806	10	US-09-745-467B-3	1	us-09-745-467B-3
4	481	6.0	1567	10	US-09-745-467B-3	1	us-09-745-467B-3
5	461	4.7	2507	9	US-09-819-106A-2	1	us-09-819-106A-2
6	456	4.6	4019	9	US-09-854-134-425	1	us-09-854-134-425
7	456	4.6	4019	9	US-09-854-134-425	1	us-09-854-134-425
8	449	4.5	2402	9	US-09-745-467B-3	1	us-09-745-467B-3
9	445	4.5	1255	9	US-09-819-106A-2	1	us-09-819-106A-2
10	445	4.5	1255	9	US-09-819-106A-2	1	us-09-819-106A-2
11	427	4.4	2412	12	US-10-124-557-5	1	us-10-124-557-5
12	427	4.4	2412	12	US-10-124-557-5	1	us-10-124-557-5
13	416	4.1	2665	10	US-09-864-761-3124B	1	us-09-864-761-3124B
14	413	4.1	2441	12	US-10-124-557-5	1	us-10-124-557-5
15	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
16	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
17	411	4.0	1404	12	US-10-124-557-5	1	us-10-124-557-5
18	410	4.0	1414	12	US-10-124-557-5	1	us-10-124-557-5
19	410	4.0	1414	12	US-10-124-557-5	1	us-10-124-557-5

[illegible][illegible]

```

07 115 TTNHDSJRKFNNSAHLKSNLSJHLEHSHSTSSNAAPRNSSTPSRND 114161174
08 279 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 114161174
09 123 FAKVYVYSIFEMANKA AFAVI KQVYVYVYVYVYVYVYVYVYVYVY 114161174
10 445 -----LALMIAOQMKKAPLJAG ----- 114161174
07 229 MUPKIPUQUPAIANQDQNSQNTFLVPTTPPTFAVAK ----- 114161174
10 465 TVQUTHSHPPHYVPSQQAQQAHTFHPQNSNPGQFPAQMSKLGQHTSHVTF 114161174
07 281 VBNKLITS -----VQSSAS ----- 114161174
10 423 LNKPSASSPSSFOQSSASSPTVNOUQOMQDPPQNNPQ HQGQCVSSSHQDNDVQ 114161174
07 416 -----FVSSQSSSSSAIPKATP ----- 114161174
10 483 QNVPTNFVVMQOQPTNQDQSSHLHPLQAKKRIHQHSAVQANIPNQGVESV 114161174
07 450 QI SQHLEHR -----ERSTQI RQIQKMLFPEREPDQ ----- 114161174
10 543 NSCAVQI QANQVQQAQQAQQAQPPQNNQVSHQPPNNQDPSLMQI HQMNNQVATISVP 114161174
07 389 QQNVGVLHG FUKKPPQ -----PQAMMASQSSQKQPPH QVAVP 114161174
10 603 QVNSNQVQOQDPSQIMQHQVPSQOQVQOQCTINQNM QSSALHG 114161174
07 446 -----RQVPSQDQVPP-SNNSQSSQ----- 114161174
10 663 VNPISQNSQSSQDQKRIHQKML SQGQHQMDP-INQMMQVQVLLQ 114161174
07 488 HQVYVQVQSLQIMVHQHQPRTVQVQPPV ----- 114161174
10 721 MCKKQVQNTQNSQVNMQPPQIMQDPTMQGPTPNQSQMMQVQVQDQMSQ 114161174
07 526 -----WAPQATP ----- 114161174
10 781 MQLQVQLRPPQPSHMQHQHQAFTANNQVSLQOMQVSSQQTMMV 114161174
07 547 -----KQV-----APHTNM-QSQ----- 114161174
10 841 ASQHSHSQSSNAFNSQANQNSQSSQDQNPVQNKVLLSHQV 114161174
07 569 -INSEMEJAVI-NVASK ----- 114161174
10 901 QVNNKQNNITANKKKKKKPKKKKNSQDINTDQIQAHLQADQHL 114161174
07 597 PKIHHQNNPPQ ----- 114161174
10 961 PKIHFSSNPQVQSSQVQDQHTQOMQQLMQVAVPQVQOQOQ 114161174
07 668 EKQQLPPT-MAMEKIRIKSEKSKRIHQSQHQPNNPPEPPIV ----- 114161174
10 1011 QQQQPPPSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSS 114161174
07 689 SQHQP-----QHTPEPQVPSQKQVAVINQV-MQSSQ- 114161174
10 1064 DQOMHQSSQSSQVAVVSLQDPA -----SVPSQDQKQMPVPT 114161174
07 738 KFAVADP-EELMKLR ----- 114161174
10 1119 QNSSSLQAMASIPASQSSQVQVQPPNNKSHVLLQNMQLQKQSSQ 114161174
07 771 EHVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174
10 1179 AITVQPP-----FVNSLSSHQ----- 114161174
07 840 SLNADPVGQGLKRPPLQSSVAT-----SQVSSP----- 114161174
10 1241 S-----TSSSLSSQFRLNASTAQIPVQVQVQVQVQVQVQVQVQVQ 114161174

```

```

07 873 SPMLQSPQNTKSPQIPQSLQALMLQVQAAASIKSPVLAASAAVHLKSPSL 114161174
10 1488 SNLITMNS-NEATPQI HKLQSVVN -----SKQSSQVAKRASPNSK- 114161174
07 944 WSSKRTIPSPQIPPNKAL-IMAS-----PAMQVNSQSSQ----- 114161174
10 1446 SSKRTIP-----SNQ-KQNSKAKELIASQVNAALQVNEHQBENQVSLP 114161174
07 985 GQSSSTVIMQVPTISQNT-SIMMSKRTIAMSSSTVY ----- 114161174
10 1488 QSTNNSSQIN QONSTVSAVAAVAVQVQVQVQVQVQVQVQVQVQVQV 114161174
07 1427 DAKFVASSDP-----DQFVASSNPSSMN ----- 114161174
10 1447 QVQVAVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174
07 1074 MPTISPKMQLQVSHSQNSQVQVQVQVQVQVQVQVQVQVQVQV 114161174
10 1507 PQQLDQVLPVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174
07 1120 MGHQSQHQPVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174
10 1565 SSSSSVAVSLP -----FVMSKPVSSSSQSSQSSQSSQSSQSSQSS 114161174
07 1180 PSNLPQSSAAVATKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174
10 1602 TSNPTTSSANSAALP----- 114161174
07 1242 PEP ----- 114161174
10 1653 PUFFTVEFNSSSLQVQKQSSQVSTHAAPLTNSQLMSSVAVVQV 114161174
07 1263 KQGLALPQMKQVQV ----- 114161174
10 1713 VQVNAISSVAVNPQIQKPLVSSRAIPVQVQVQVQVQVQVQVQVQV 114161174
07 1329 PH-----HRMSIAQSSQVQVQVQVQVQVQVQVQVQVQVQV 114161174
10 1773 PQVNSAQNQDQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQ 114161174
07 1468 ----- 114161174
10 1828 ELKQKQVQSSLEKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174
07 1403 KQMAVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174
10 1888 KQSSVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 114161174

```

RESULT 3
 1. Sequence 56, Application US/09919497
 2. Patent No. US2002010662A1
 3. GENERAL INFORMATION:
 4. APPLICANT: Mottet, George L.
 5. TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOCRINAL CANCER
 6. CURRENT APPLICATION NUMBER: US/09919497
 7. CURRENT FILING DATE: 2001-07-31
 8. PRIOR APPLICATION NUMBER: US 60/221,735
 9. PRIOR FILING DATE: 2000-07-31
 10. NUMBER OF SEQ ID NOS: 100
 11. SOFTWARE: Patent In version 3.0
 12. SEQ ID NO 56
 13. LENGTH: 1806
 14. TYPE: PRT
 15. ORGANISM: Homo sapiens
 16. FEATURE:
 17. NAME/KEY: UNSURE
 18. LOCATION: (758) ..(759)
 19. OTHER INFORMATION: X44 - any amino acid
 20. NAME/KEY: UNSURE
 21. LOCATION: (803) ..(807)

DB 1204 SCLASNTLTSNPPQJTRP11VWYSSQJCTPRSTC1DLEVLVYNONEVVAHED 1204
 QY 345 PVSSTPPTLAPNPAISQJLPHREBSLQJLKDLOMPLPEJKEP JGASSTN 345
 DB 1204 TPRDTSVYSSQJATPPRDLSPDTRSSMI --- KVLAMNLOJLOJAAJN09 1414
 QY 404 VLDQKPPRDLLOAMMOSQJSGQJGP - KJWVAFQPOJH0V9 SPLEJVSFNS 402
 DB 1415 ALGQJWBPQJCT --- SCL PPRDQJSLDLSRVSISAQJ - PVLJ SLPJL FNSQJ 1406
 QY 403 QSGTJ --- --- GQJQJHJHJLAW 1472
 DB 1407 SFGJSLAHVJALQJPRVSTSPVASSNSIMHJSGQJPSVSQLPQVJVSJVLJN1 1426
 QY 478 LKJQJVEEKKKQJCT - QVVVQJQJ SLOJMMVJHJHJGQJGP - QJQJ SCLMAJ 3700
 DB 1426 VNMVQJLPEKJLQJQJLHJLJLQJQJ QJQJKTALQJQJSGJSLAVJHJ - QJQJQJQJLJN 1402
 QY 541 LF --- PPSQJLNMHJSL --- JH --- GMAJH --- 552
 DB 1403 VNOATKJQJQJYVJNJSVAPJQJQJAVJHJQJQJQJYVJHJASMBRQJQJQJQJ 1442
 QY 554 --- JN --- MJSQJQJLQJQJAMT --- --- NSHJQJQJ - NVJN --- 561
 DB 1543 SHQJQJQJQJLQJQJQJQJQJQJQJQJQJQJQJQJQJQJQJQJQJQJQJ 1462
 QY 582 ASKQJ 609
 DB 1603 VQJ 1662
 QY 620 - RFPJQJ 677
 DB 1663 MRQJ - QJQJLQJQJQJ --- LQVJHJQJQJQJQJQJQJQJQJQJQJQJQJ 1709
 QY 678 PRJVEJQJ 729
 DB 1710 QJPS - FQJLQJQJQJ --- VJSEJTSJQJQJQJQJQJQJQJQJQJQJQJQJ 1763
 QY 721 --- --- JNV --- --- NMSSNQJQJQJQJQJQJQJQJQJQJQJQJQJ 764
 DB 1761 QVQJ 1817
 QY 754 QJ 799
 DB 1818 KLQJ 1877
 QY 800 LREJLQJ 859
 DB 1878 KJQJ 1916
 QY 860 JNPJQJ 909
 DB 1917 JNQRKJQJ 1971
 QY 910 VJSSAAJQJ 964
 DB 1972 VJSSJLQJ 2004
 QY 962 JQVJQJ 2012
 DB 2025 NSNJQJ 2062
 QY 1013 MSKQJ 2068
 DB 2083 QJ 2132
 QY 1059 JNQJ 2136
 DB 2133 QJ 2192
 QY 1119 JNQRKJQJ 2197
 DB 2183 JNQRKJQJ 2207

QY 1176 JNQRKJQJ 2214
 DB 2248 JNQRKJQJ 2247
 QY 1215 JNQRKJQJ 1241
 DB 2298 JNQRKJQJ 1241
 QY 1262 QJ 1421
 DB 2401 JNQRKJQJ 2400
 QY 1422 --- --- --- --- 1422
 DB 2407 QJ 2400
 QY 1422 --- --- --- --- 1422
 DB 2407 QJ 2400
 QY 1468 JNQRKJQJ 1466
 DB 2517 JNQRKJQJ 2576
 QY 1407 ADVQJ 1426
 DB 2577 JNQRKJQJ 2579
 RESULT 7
 US-09-915-543-15
 Sequence 425, Applicant US-09-915-543-15
 Patent No. US2002011056A1
 GENERAL INFORMATION
 APPLICANT: Resed, Steven G.
 APPLICANT: Henderson, Robert A.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Elmer, Steven P.
 APPLICANT: Melamita, Kenneth
 APPLICANT: Alford, Paul A.
 APPLICANT: Societ, Heather
 APPLICANT: Indrias, Carol Yoseph
 APPLICANT: Henderson, David R.
 APPLICANT: Elmer, Mark
 APPLICANT: Malton, Ade
 APPLICANT: Kallou, Michael D.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
 FILE REFERENCE: 210121-47509
 CURRENT APPLICATION NUMBER: US-09-915-543-15
 CURRENT FILING DATE: 2000-12-14
 NUMBER OF SEQ ID NOS: 587
 SOFTWARE: PostScript Windows Version 4.0
 SEQ ID NO 425
 LENGTH: 4019
 TYPE: DKT
 ORGANISM: Homo sapiens
 US-09-915-543-15
 GENE MATCH
 best local similarity: 4.0% Score 406.5, 100.10% Identity 4019
 Matches: 485 Conserved 1792 Missed 692, Indels 687, Gaps 90
 QY 17 QJ 17
 DB 811 QJ 17
 QY 77 JNQRKJQJ 17
 DB 807 JNQRKJQJ 17
 QY 198 JNQRKJQJ 17


```

Query Match 4.68 Score 45.57 100.00 Length 1.00
Host Local Similarity 21.48 Prod. No. 110-008
Matches 261: Conserved 125: Mismatches 400: Indels 147: Gaps 653

009 SNNKRRSTALINQIS-----ALRNDKPLVQVPAVANDQVNSQNTSLITIT 261
114 SNNQIKRIGILSVQIEKCAINIALA-----GNSQVGRN 150
009 AVALKPAIPKIDRESEVYINKLIPVSPASSIT----- 603
151 -----GPM-----VANSVMEKATPMASSVILKNNIAVILTEPMAVSSSR 196
009 -----GNSITNNKAVTVVQVNSSSAP-----KAPPEVSSDEPIL 345
196 APENPPIQPEKPRIASO SIQMIPLSLHIGQSHSPSLAFPHM QPVSVN 249
009 446 E--NIGLISQILERHRSQITLQIQMIPEDEKELQASRIQVZNVN--GKRP 403
250 KQNNANPQIQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 600
009 404 GPTQAMMAVSQILKQGPRIQVAPPEQVQVQVQVQVQVQVQVQVQVQVQVQV 459
009 401 GTRQPTAPV-----VP-----VPPNNQPRQALQV 641
009 460 -VALLHMTPLQIAIK-----LQPEVEERKRPVQVQVQVQVQVQVQVQVQV 612
009 412 QSLQIMLANQ--GKKKALPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 471
009 513 GPPVYQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 570
009 472 -PEPVY-----VQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 406
009 571 SEMEQNVPINVASPPISTVSWPQVQVQVQVQVQVQVQVQVQVQVQVQVQV 609
009 406 SLQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 459
009 620 REPVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 675
009 460 NNPLQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 490
009 676 -LEPPIVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 627
009 491 VMDQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 545
009 728 SNNQITLQKMAA--GAG--PEEMKLPVQVQVQVQVQVQVQVQVQVQVQVQV 760
009 546 APVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 593
009 781 REPPIQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 826
009 600 QVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 652
009 827 -----NLSINAPVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 674
009 658 QVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 714
009 877 -----GNSINIKPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 932
009 715 QVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 772
009 943 WFSST-----KPPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 981
009 773 VNNSSQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 814
009 982 NIPNLSSTTVTAPPEPILSQNTLSIMSPKSTAPNSTPQVQVQVQVQVQV 1041
009 815 S-----QNPVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 841
009 1042 -----ARSDNIPNNNIPQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1008
009 842 HVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 808
009 1099 GPNTIPVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1138

```

```

1149 -----QVPTIRINQIS-----GNSPFG GMAVQVQVQVQVQVQVQVQV 1153
009 924 KNSQDILNTPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 980
009 1198 KQNSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1257
009 981 -----ETITALQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1119
009 1258 PQRQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1268
009 1020 PQRNSRDEE 1030

RESULT 9
US-09-819-104A-5
Sequence 5: Application US/09810104A
Publication No. US20040271471
GENERAL INFORMATION:
APPLICANT: Ciba, J. Ltd.
TITLE OF INVENTION: NOVEL NUTRIENT SUPPLEMENT FOR CEREAL GRAIN
TITLE OF INVENTION AND USES THEREOF
FILE REFERENCE: US/040104A
CURRENT APPLICATION NUMBER: US/040104A
PRIOR APPLICATION NUMBER: 60/194,138
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ. ID NOS.: 6
SOFTWARE: Patented Ver. 2.0
SEQ. ID NO. 5
LENGTH: 2462
TYPE: PKI
ORGANISM: MUS musculus
US-09-819-104A-5

Query Match 4.58 Score 44.97 100.00 Length 2462
Host Local Similarity 20.69 Prod. No. 428-008
Matches 454: Conserved 153: Mismatches 646: Indels 552: Gaps 621

009 4 SNNKRRSTALINQIS-----ALRNDKPLVQVPAVANDQVNSQNTSLITIT 261
009 747 SVPSRSEFALRIDQPKTQIEALPAALQVQVQVQVQVQVQVQVQVQVQVQV 792
009 64 QVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 122
009 793 VPIASQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 803
009 123 KQNS--AMIRKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 101
009 854 KQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 879
009 182 LEMANKAFAVIRKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 924
009 880 LEAFVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 924
009 925 VDEPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 975
009 299 PPIQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 976
009 976 -----VIRVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1007
009 459 RERSTQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 416
009 1008 -----DVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1046
009 417 GRQV-----RQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 448
009 1049 FLSHAGKRLSHQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1109

```


[illegible][illegible]

05 562 -----LPGAV--MINSEMRN-----VENVANSN-ILSVSWIT-509
 06 423 PGDPCANLPLAABERATLISB-PAVING-IGGEKPAIT-KCARPAPACIOAAT-704V 482
 07 597 PRIGGRTFEGGCFESGCGEPFINDCOLSEEPQOQAEKGLPLP-MAH-705M 656
 08 483 PGDGMKMGPSPI-----GIVSDK-PPSPGSGSGRIG-----PPDS-706V 440
 09 657 EKNMILNSU-REHEPNN-LEPR-IVERT-SSK-APK-IT-707M 715
 10 441 ---MOTFPGKRNAPAKN-----DERG-PODPPGPG-GRN-717 471
 11 719 MCOLVNLVNSNSOM-POKME-AVAGPEEM-LEK-NDSSM-PAV-KWET-719 472
 12 472 PGDGTGGRKOD-OPR-POU-GR-719-----CHVNDKSG-----719K-719A 520
 13 774 POGYVMPHRE-LEMSQ-SSNS-LEK-RE-719-GR-719-719-719-719 442
 14 521 PGABKIDAV-APG-GR-719-719-719-719-719-719-719-719 728
 15 444 LAFV-GR-719-719-719-719-719-719-719-719-719 888
 16 579 GMP-GR-719-719-719-719-719-719-719-719-719 902
 17 886 POTS-GR-719-719-719-719-719-719-719-719-719 909
 18 613 PGDGTGGRKOD-OPR-POU-GR-719-----CHVNDKSG-----719K-719A 520
 19 940 PLOSP-GR-719-719-719-719-719-719-719-719-719 902
 20 661 PAFV-GR-719-719-719-719-719-719-719-719-719 902
 21 984 LPS-GR-719-719-719-719-719-719-719-719-719 902
 22 721 GPM-GR-719-719-719-719-719-719-719-719-719 902
 23 1043 KSN-GR-719-719-719-719-719-719-719-719-719 902
 24 748 PMS-GR-719-719-719-719-719-719-719-719-719 902
 25 1100 PNF-GR-719-719-719-719-719-719-719-719-719 902
 26 793 P-GR-719-719-719-719-719-719-719-719-719 902
 27 1159 SG-GR-719-719-719-719-719-719-719-719-719 902
 28 827 QV-GR-719-719-719-719-719-719-719-719-719 902
 29 1216 TOP-GR-719-719-719-719-719-719-719-719-719 902
 30 879 GSI-GR-719-719-719-719-719-719-719-719-719 902
 31 1271 SHM-GR-719-719-719-719-719-719-719-719-719 902
 32 911 AKS-GR-719-719-719-719-719-719-719-719-719 902
 33 1314 NMR-GR-719-719-719-719-719-719-719-719-719 902
 34 967 NVC-GR-719-719-719-719-719-719-719-719-719 902
 35 1370 THG-GR-719-719-719-719-719-719-719-719-719 902
 36 1009 -GR-719-719-719-719-719-719-719-719-719 902

05 562 -----LPGAV--MINSEMRN-----VENVANSN-ILSVSWIT-509
 06 423 PGDPCANLPLAABERATLISB-PAVING-IGGEKPAIT-KCARPAPACIOAAT-704V 482
 07 597 PRIGGRTFEGGCFESGCGEPFINDCOLSEEPQOQAEKGLPLP-MAH-705M 656
 08 483 PGDGMKMGPSPI-----GIVSDK-PPSPGSGSGRIG-----PPDS-706V 440
 09 657 EKNMILNSU-REHEPNN-LEPR-IVERT-SSK-APK-IT-707M 715
 10 441 ---MOTFPGKRNAPAKN-----DERG-PODPPGPG-GRN-717 471
 11 719 MCOLVNLVNSNSOM-POKME-AVAGPEEM-LEK-NDSSM-PAV-KWET-719 472
 12 472 PGDGTGGRKOD-OPR-POU-GR-719-----CHVNDKSG-----719K-719A 520
 13 774 POGYVMPHRE-LEMSQ-SSNS-LEK-RE-719-GR-719-719-719-719 442
 14 521 PGABKIDAV-APG-GR-719-719-719-719-719-719-719-719 728
 15 444 LAFV-GR-719-719-719-719-719-719-719-719-719 888
 16 579 GMP-GR-719-719-719-719-719-719-719-719-719 902
 17 886 POTS-GR-719-719-719-719-719-719-719-719-719 909
 18 613 PGDGTGGRKOD-OPR-POU-GR-719-----CHVNDKSG-----719K-719A 520
 19 940 PLOSP-GR-719-719-719-719-719-719-719-719-719 902
 20 661 PAFV-GR-719-719-719-719-719-719-719-719-719 902
 21 984 LPS-GR-719-719-719-719-719-719-719-719-719 902
 22 721 GPM-GR-719-719-719-719-719-719-719-719-719 902
 23 1043 KSN-GR-719-719-719-719-719-719-719-719-719 902
 24 748 PMS-GR-719-719-719-719-719-719-719-719-719 902
 25 1100 PNF-GR-719-719-719-719-719-719-719-719-719 902
 26 793 P-GR-719-719-719-719-719-719-719-719-719 902
 27 1159 SG-GR-719-719-719-719-719-719-719-719-719 902
 28 827 QV-GR-719-719-719-719-719-719-719-719-719 902
 29 1216 TOP-GR-719-719-719-719-719-719-719-719-719 902
 30 879 GSI-GR-719-719-719-719-719-719-719-719-719 902
 31 1271 SHM-GR-719-719-719-719-719-719-719-719-719 902
 32 911 AKS-GR-719-719-719-719-719-719-719-719-719 902
 33 1314 NMR-GR-719-719-719-719-719-719-719-719-719 902
 34 967 NVC-GR-719-719-719-719-719-719-719-719-719 902
 35 1370 THG-GR-719-719-719-719-719-719-719-719-719 902
 36 1009 -GR-719-719-719-719-719-719-719-719-719 902

05 562 -----LPGAV--MINSEMRN-----VENVANSN-ILSVSWIT-509
 06 423 PGDPCANLPLAABERATLISB-PAVING-IGGEKPAIT-KCARPAPACIOAAT-704V 482
 07 597 PRIGGRTFEGGCFESGCGEPFINDCOLSEEPQOQAEKGLPLP-MAH-705M 656
 08 483 PGDGMKMGPSPI-----GIVSDK-PPSPGSGSGRIG-----PPDS-706V 440
 09 657 EKNMILNSU-REHEPNN-LEPR-IVERT-SSK-APK-IT-707M 715
 10 441 ---MOTFPGKRNAPAKN-----DERG-PODPPGPG-GRN-717 471
 11 719 MCOLVNLVNSNSOM-POKME-AVAGPEEM-LEK-NDSSM-PAV-KWET-719 472
 12 472 PGDGTGGRKOD-OPR-POU-GR-719-----CHVNDKSG-----719K-719A 520
 13 774 POGYVMPHRE-LEMSQ-SSNS-LEK-RE-719-GR-719-719-719-719 442
 14 521 PGABKIDAV-APG-GR-719-719-719-719-719-719-719-719 728
 15 444 LAFV-GR-719-719-719-719-719-719-719-719-719 888
 16 579 GMP-GR-719-719-719-719-719-719-719-719-719 902
 17 886 POTS-GR-719-719-719-719-719-719-719-719-719 909
 18 613 PGDGTGGRKOD-OPR-POU-GR-719-----CHVNDKSG-----719K-719A 520
 19 940 PLOSP-GR-719-719-719-719-719-719-719-719-719 902
 20 661 PAFV-GR-719-719-719-719-719-719-719-719-719 902
 21 984 LPS-GR-719-719-719-719-719-719-719-719-719 902
 22 721 GPM-GR-719-719-719-719-719-719-719-719-719 902
 23 1043 KSN-GR-719-719-719-719-719-719-719-719-719 902
 24 748 PMS-GR-719-719-719-719-719-719-719-719-719 902
 25 1100 PNF-GR-719-719-719-719-719-719-719-719-719 902
 26 793 P-GR-719-719-719-719-719-719-719-719-719 902
 27 1159 SG-GR-719-719-719-719-719-719-719-719-719 902
 28 827 QV-GR-719-719-719-719-719-719-719-719-719 902
 29 1216 TOP-GR-719-719-719-719-719-719-719-719-719 902
 30 879 GSI-GR-719-719-719-719-719-719-719-719-719 902
 31 1271 SHM-GR-719-719-719-719-719-719-719-719-719 902
 32 911 AKS-GR-719-719-719-719-719-719-719-719-719 902
 33 1314 NMR-GR-719-719-719-719-719-719-719-719-719 902
 34 967 NVC-GR-719-719-719-719-719-719-719-719-719 902
 35 1370 THG-GR-719-719-719-719-719-719-719-719-719 902
 36 1009 -GR-719-719-719-719-719-719-719-719-719 902

07 808 GRINSLIMPLPLNNSSNPTSLNTPVYRGJAKRPLJL SVAGS.VINSE...NF KSPJ 807
 10 554 GPKCKGAGGPGGAGGAPL-DLQMPGPKG...-S...K... 505
 07 808 MHVUSPMI GSPST...-NLKSPDTPSQA...LAPAAANK...-S...VLSSAA 919
 10 596 - - - - -GPGGPGAAVYCKGCKGPGGPGPGP- PGVAGGPGDK D...H...G...A... 444
 07 917 SPVHLKSPSLAVSPG--WLSSTKPLPSTGTPENIKAPL KASPAW...N...V... 971
 10 645 - - - - -PRSGPGGPGGPGGPGGPGGPGGAGNUNEPGCKGCKG...K...H...G...A... 604
 07 972 -PLTASPA - - - - -SVNLPLSSSTPYIMPEPLLSQNPUS...MSK...K...A... 117
 10 695 GPGGSGGPGGPGGPGGCKGCKGSGGPGGAGPGVAGLPPGPGSN... 742
 07 1018 MPSTPLPYHAKLIVANSISGSPASPNLSNNNMIMKINTI- UNP...S...N...V...M... 1074
 10 743 - - - - -GPGGSGS--PGKGGPGGAGNIGVAGS...V...S...K...H...G...A... 780
 07 1075 PLTSPMOMIQLSHSNOMPSNAPVNTIPHVPMEGLMSHMTM-PLSSGTPVAG... 1144
 10 781 P - - - - -GPKSGPGAGP - - - - -PGVAGPGAGNIGAGAGL...L...M... 818
 07 1134 RGPVGGPGPVVSPGCGVTPHNMFGC--GQSGFPGMGPGVGE-GPGGRLN...P...S...A... 1190
 10 819 - - - - -PGVSGVCKGCKPGVANTLSGCKGPGG PGV...L...A...H... 859
 07 1191 ALGPKVAGGPGSTVIGNSMPVETP--DLOVTPGAGTGPFGSLK...S...K... 1246
 10 860 G--E...H...D...N...S...-D...L...P...R...S...P...K...C...K...G...S...G...A... 899
 07 1247 QLLQVPRHVPKCKGCKGPG--GPHESHMOMGELGAMHMLALPG--M...H...V...L... 901
 10 897 - - - - -APGHPGHPVAGP--APKSGDKGSGSLAGIAGAPG...N...S...A...H... 941
 07 1402 PLCT - - - - -ALSPGCH--NMPD--P...L...G...M...M...P...H...M...S...A...S...H... 1444
 10 942 -PGGCKGELGEGVAGLKHGKPGGNGAGLPGHAP--GCGALG--S...P...A...H... 992
 07 1445 PLTSMNAAVMTGPGKPGVAGLTYTHGIVVSP...M...M...S...G...M...G...N...M...H...G...H... 1404
 10 993 PGTGSPG-- - - - - -PGKD-GLSG--H...G...L...G...P...G... - - - - -PRG... 1079
 07 1405 MAADVMOGPGSGGAGNG 1422
 10 1020 NRGKGSRLSPGHPGPG 1047
 RESULT 9
 US-08-175-155-68
 Sequence 68, Application US/08175155
 Patent No. 5611644
 GENERAL INFORMATION:
 APPLICANT: Ferrari, Franco A.
 APPLICANT: Cappello, Joseph
 APPLICANT: Cissman, John W.
 APPLICANT: Dorman, Mary A.
 TITLE OF INVENTION: Methods for Preparing Synthetic
 NUMBER OF SEQUENCES: 69
 CORRESPONDENT ADDRESS:
 ADDRESS: Fitch, Hothorn, Post, Attribution & Invention
 STREET: Four Emburgado Court, Suite 400
 CITY: San Francisco
 STATE: CA
 COUNTRY: US
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIA TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: pc-tos/ms-dos
 SOFTWARE: Patent to Release #1.0, Version #1.00

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/175-155
 FILING DATE: 29-08-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rowland, Jeffrey T.
 REGISTRATION NUMBER: 20015
 REFERENCE/DOCKET NUMBER: A-55186 5/HIK
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-781-1989
 TELEFAX: 415-498-3249
 INFORMATION FOR SEQ ID NO: 68:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 837 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MODIFIED TYPE: peptide
 US-08-175-155-68
 Query Match: 6,781 Score 515.57 DB 1: Local 8412
 Best Local Similarity: 26.08; Prod. No. 10-231
 Matches: 311; Conservative: 507; Mismatches: 499; Indels: 447; Gaps: 74
 275 PRESQVE -NKLISVSPASSTPLPGKI GPNSTNNKAV--TIVSNGSSSARLK 429
 10 DNNINPVTLNKL AAHPPASNDMAAGTLPFGVGLSVAWVATLPGVGLNSGAGT 67
 07 430 AHPD--PWSSTPEPTLGNPKLSSGQLEHREKSTGLKTDQNMPEKKEPGASST 487
 10 68 TPGVGLNSGAGTLPPT--PGGL-----PSSVAGG 443 100
 07 488 PGNKAV - - - - -LDDGKKPKPSTLAAVMASSQSLKGVKPKVAVAPVAGVH... 442
 10 101 PGTGSLSPAGTLPVGLGTPSGVAGTLPVGLGSGVCA - - - - -GVGQDPSGVAF 194
 07 443 DDAVPSMNSGSLGPHLDHMTPEGLAWKLGGVTFYERKCKELVVVG...SLGIMV 502
 10 157 GTPVGLNSGAGT-- - - - - -TP----- 174
 07 503 HGHGPGVAVP...PPYUMPSRGMWAGTLPESNGINMHSPLPGMAHNMPSGML 502
 10 175 - - - - -GPGGLGSGAG--GTPGGLGPG--SRVAPLPGVGLP--GSGVAGTLPVGL 244
 07 564 PGPAGMINSMPGVPVND--ASRQSLGVSVAELGVKTPGKNTLPVGLTSPGK 919
 10 225 PGGV - - - - -APGTPGPGVGLNSGAGTLPVGLNSGAGTLPVGLNSGAGT 276
 07 620 KPPNGVGLSTMPVGLAKGVLTPGAMGCKLPSMNNKNTLSSGGMGPGVANNITP 679
 10 277 GTPGGLNSG-- - - - - -PG-----AGTTPGGLNSG-- - - - - -PG 604
 07 680 LVEPSLSKPKD...KGLTPGNG-- - - - - -PAREEPGVMTSGGVN...M...N...S...G...M... 745
 10 405 - - - - -APGTPG--PGVGLNSGAGTLP-- - - - - -PGVGLNSG-- - - - - - 442
 07 746 KRRVAGATHEEMKLPVGSMLPAQGRVVP...P...G...V...G...V...M...K...P...S... 794
 10 444 - - - - -PGV-----PGV-----GTPGGLGSGVAGTLPVGLNSG-- 602
 07 795 SGLNKLKEGPGVPTSLSIMPLPLTINSNPSTLNTAF - - - - -LV...G...K...P...L...S... 849
 10 44 - - - - -GAGTPGPGVGLNSGAVNP-- - - - - -GVGGLNSGAGTLPVGLNSGAGT 2412
 07 850 VAGSG--VHSKNTNLKSLTHVUSPMIGSPSLNKSGVTSVATMLAPAAANKS 907
 10 433 TPGVGLNSGAGTLPPT-- - - - - -GTPGSG-- - - - - -AGTTPG--G...S...N...G...A...H... 441
 07 908 PVLSSAAASVHLKSPSLAV-- - - - - -SPGWSSTKPKPLGSGTLPENIKAPL KASPAW 902
 10 462 QGLPGSVCA - - - - -DGLPGVGLNSGAGTLPPT-- - - - - -G...L...S...N...G...A...H... 519
 07 904 GNVSGGPTPTASLFAVNPVGLSSSTPTTPGPPPLNSNPUS...M...S...K...T...A...M...S... 1022

